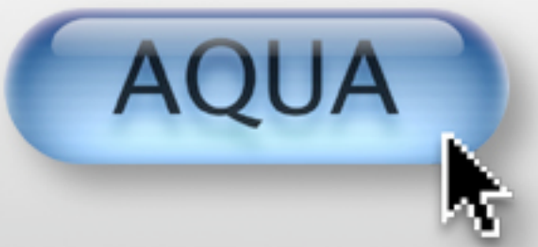
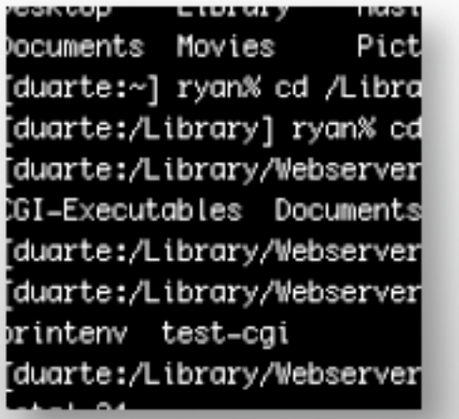
# Search Kit

## A Powerful Text Searching Framework

Session 422





# Search Kit

## A Powerful Text Searching Framework

**Wayne Loofbourrow**
Engineering Manager, Info Apps
wrl@apple.com

# Introduction

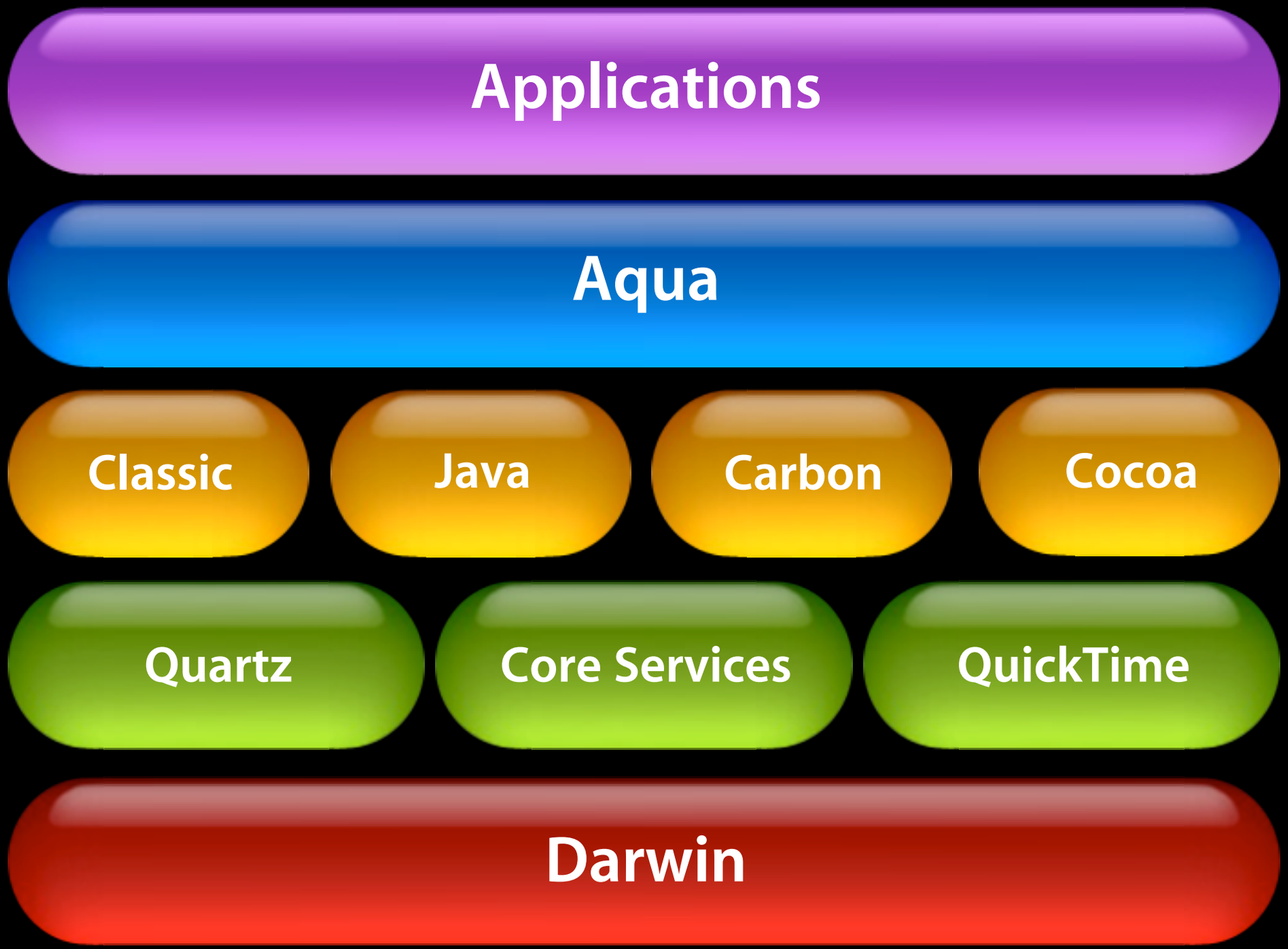
## A powerful text searching framework

- How it's used in Mac OS X
- What it can do
- How to add search to your application

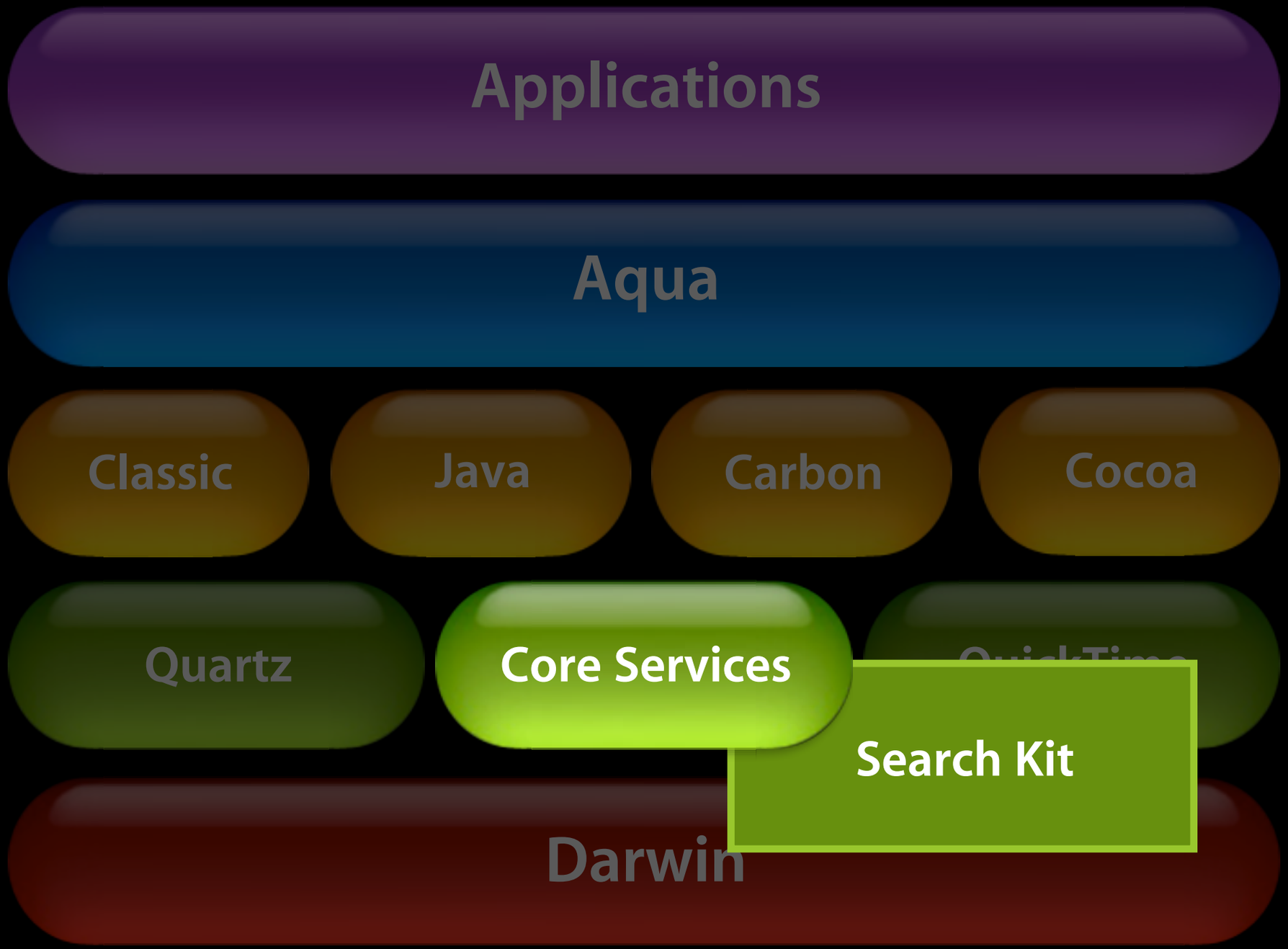
# What You'll Learn

- What is Search Kit?
- Challenges
- Indexing in 3 Steps
- Searching in 3 Steps
- Summarization

# Technology Framework



# Technology Framework



# History

- SearchKit is the Mac OS X version of V-Twin or AIAT

# History

## Mac OS 9

- Apple Information Access Toolkit (AIAT)
- C++ API
- Custom data types
- Complex API

# History
## Mac OS X

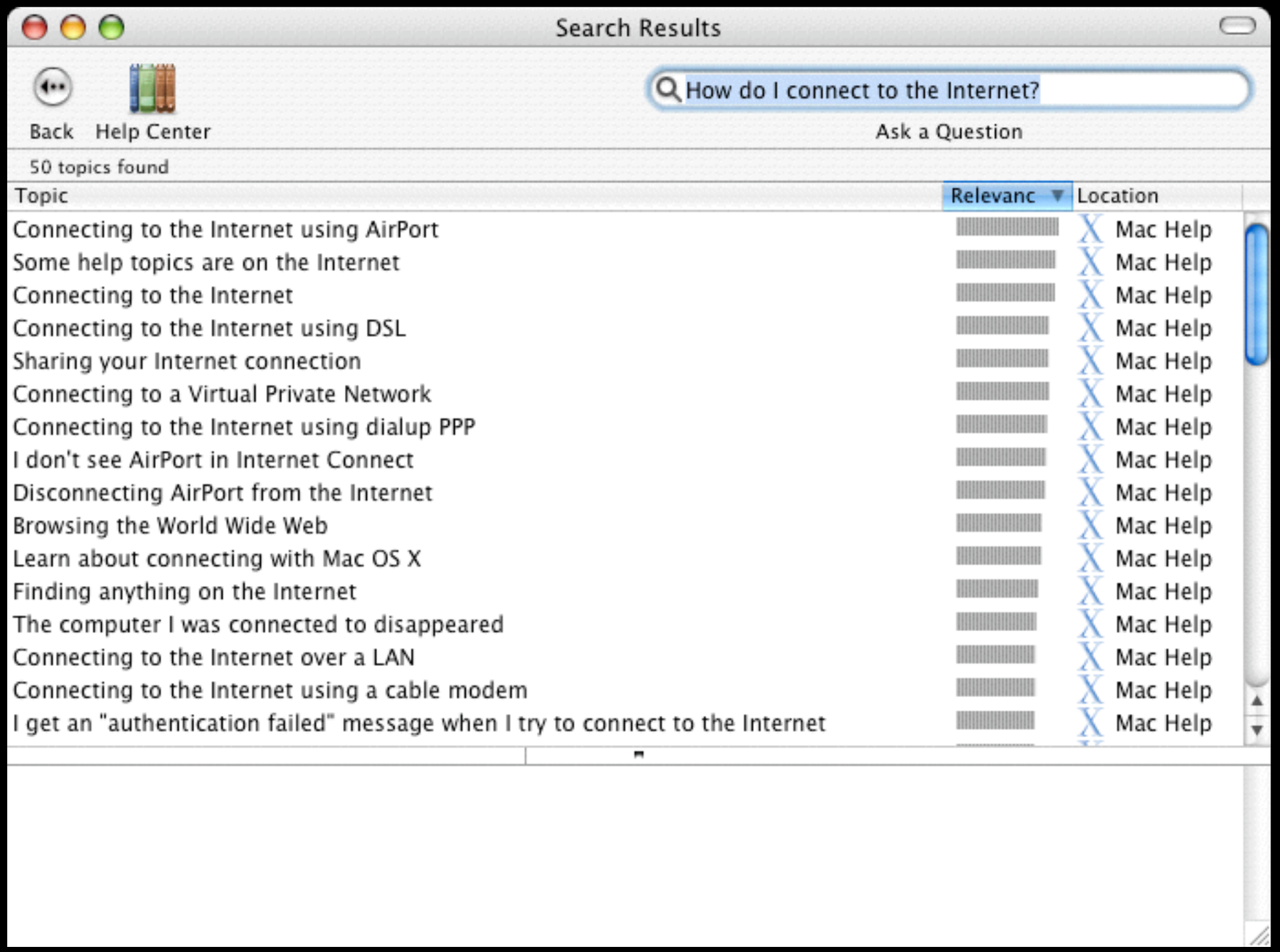
- SearchKit
- C API
- CF types
- Streamlined API

# Search Kit In Action

## Address Book

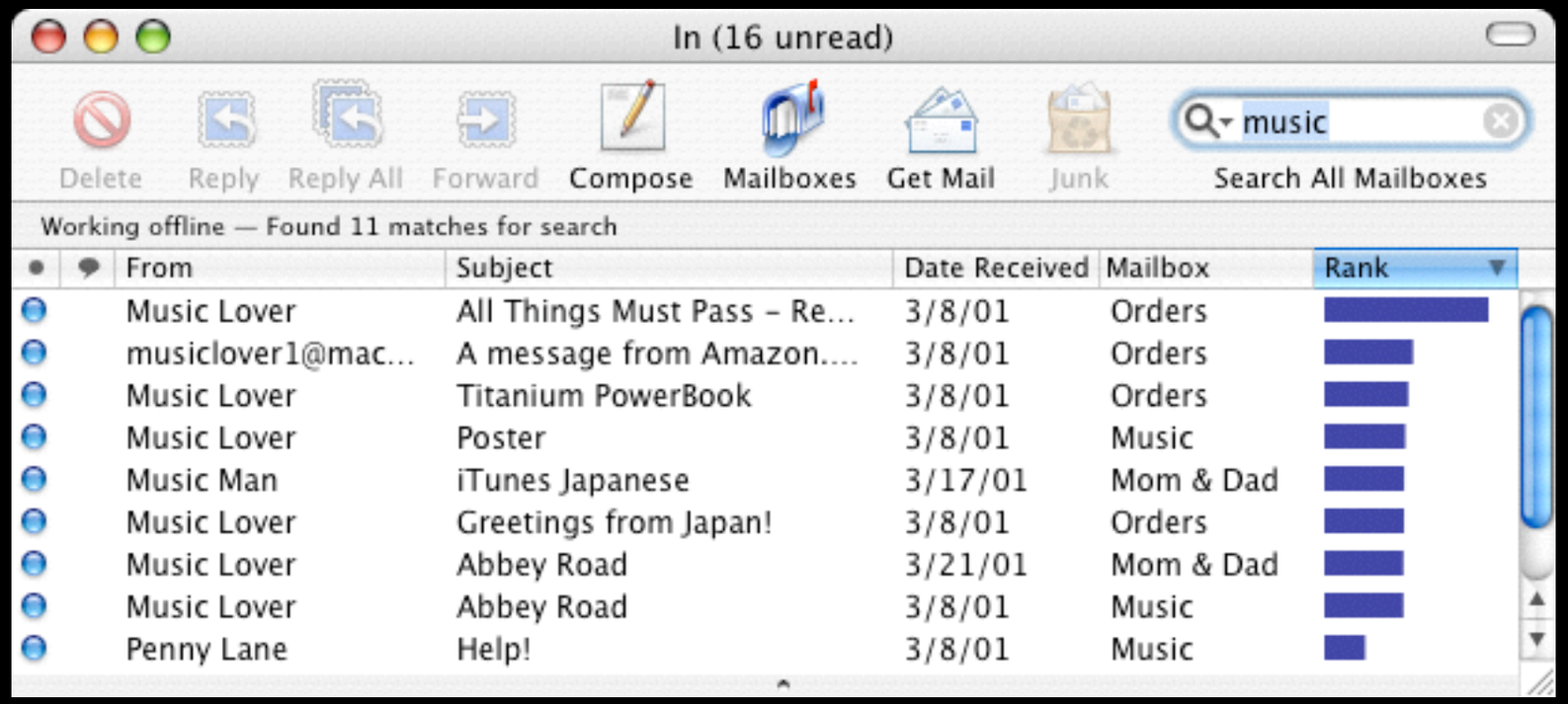
# Search Kit In Action

## Help



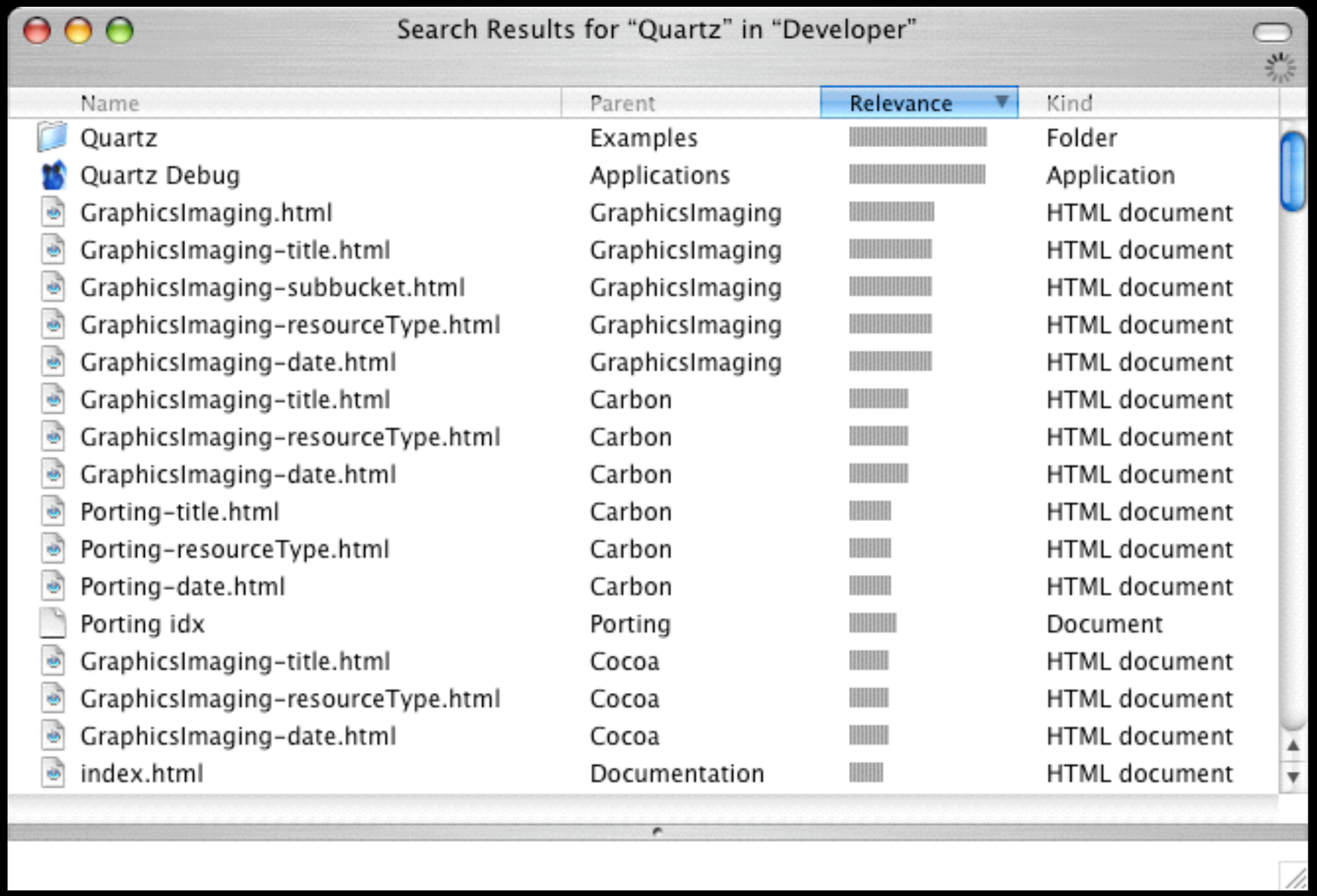
# Search Kit In Action

## Mail



# Search Kit In Action

## Finder



# Demo

**David Cásseres**
Find by Content Tech Lead

# Search Kit

## Challenges

- Relevance
- Languages
- Flexibility
  - Documents
  - Queries

# Search Kit

## Relevance

**How do I connect to the Internet?**

- What documents match?

# Search Kit

## Relevance

How | do | I | connect | to | the | Internet

- What documents match?

# Search Kit

## Relevance

How | do | I | connect | to | the | Internet

- What documents match?
- Are documents containing

do | I | to | the

# Search Kit

## Relevance

How do I connect to the Internet

- What documents match?
- Are documents containing
  - do I to the
- As relevant as those containing

  How connect Internet

# Search Kit

## Languages

- **The quick brown fox jumps.**

- Splitting into words

# Search Kit

## Languages

速い茶色のフォックスは跳ぶ。

- Splitting into words

# Search Kit

## Languages

速い茶色のフォックスは跳ぶ。

- Splitting into words
- Japanese Language Analysis

# Search Kit

## Languages

速 | い | 茶色 | の | フォックス | は | 跳ぶ

- Splitting into words
- Japanese Language Analysis

# Search Kit

## Flexibility

- What is a document?

# Search Kit

## Flexibility

- What is a document?

# Search Kit

## Flexibility

- What is a document?

# Search Kit

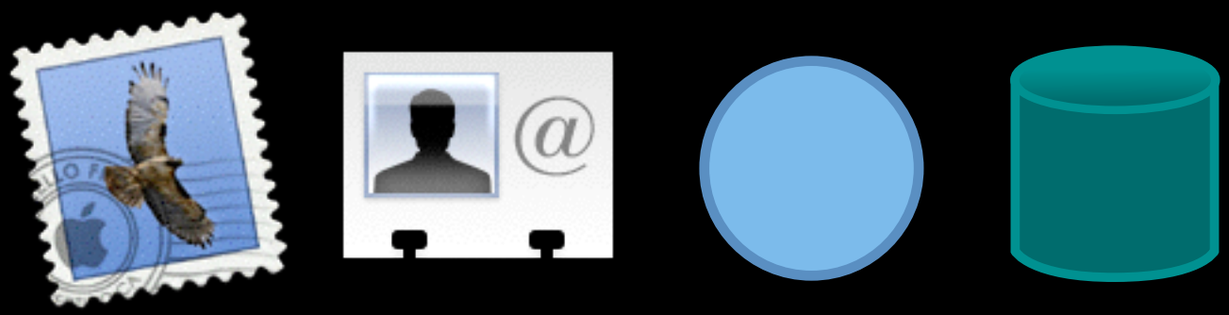
## Flexibility

- What is a document?

# Search Kit

## Flexibility

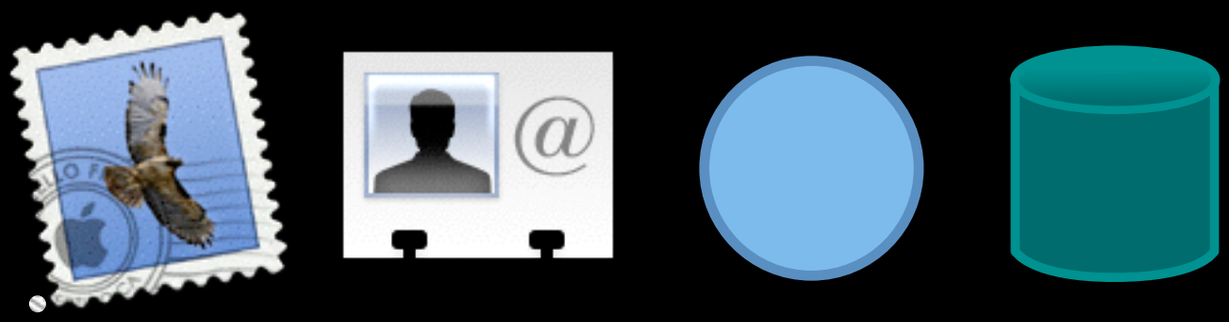
- What is a document?



# Search Kit

## Flexibility

- What is a document?



- Anything that contains text

# Search Kit

## Flexibility

- What is a query?

# Search Kit

## Flexibility

- What is a query?

**How do I connect to the Internet?**

# Search Kit

## Flexibility

- What is a query?

**How do I connect to the Internet?**

**San Fran Fri Jun 27 WWD|**

# Search Kit

## Flexibility

- What is a query?

**How do I connect to the Internet?**

**San Fran Fri Jun 27 WWD|**

**Strong & (Mad | Bad | Sad)**

# Search Kit
## Flexibility

- What is a query?

**How do I connect to the Internet?**

**San Fran Fri Jun 27 WWD|**

**Strong & (Mad | Bad | Sad)**

**chocolate +sundae -nuts**

# Search Kit

## Flexibility

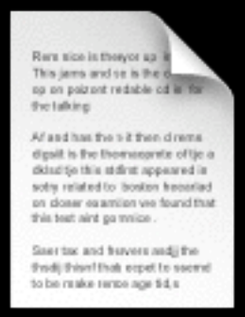
- What is a query?

**How do I connect to the Internet?**

**San Fran Fri Jun 27 WWD|**

**Strong & (Mad | Bad | Sad)**

**chocolate +sundae -nuts**



# Typical Usage Scenario

- Index documents
- Search the index

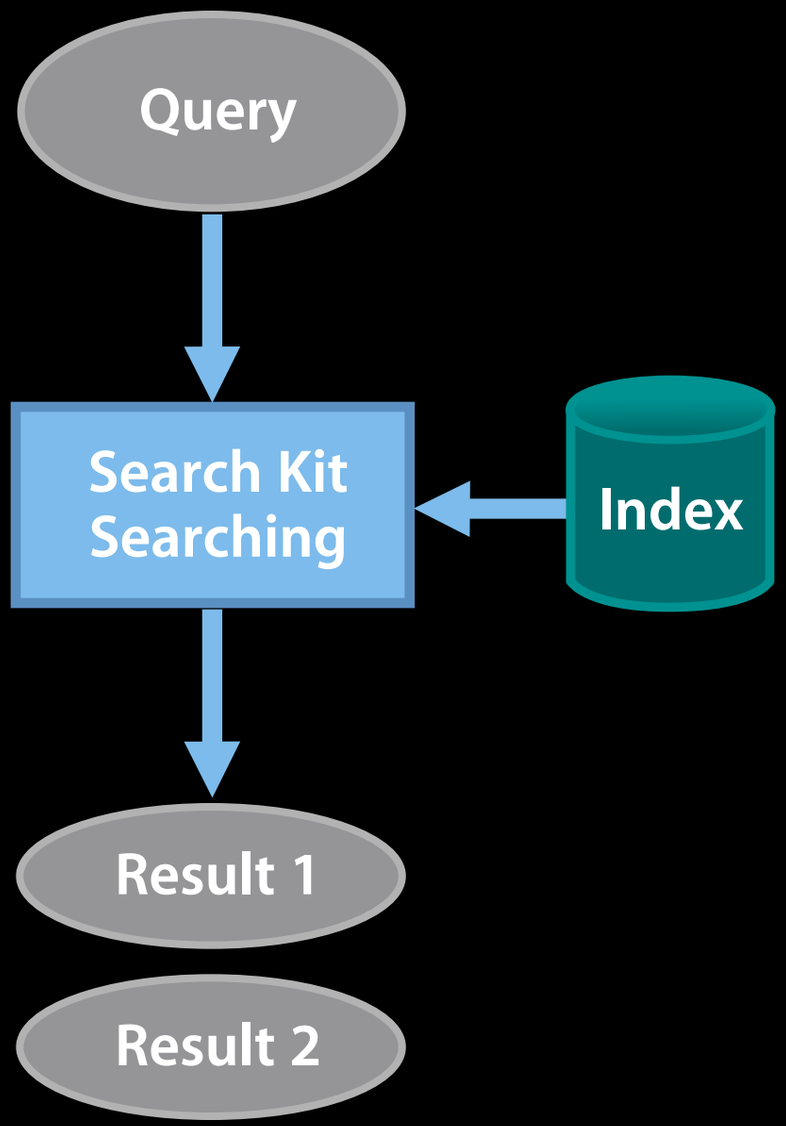
# About the Index

- Purpose
  - Makes searching fast
  - Don’t need original docs
- Stores
  - Document references
  - Info about words in each
- Where is index?
  - In memory
  - In a file

# Indexing

- Adding documents
- Search Kit
  - Analyzes the text
  - Updates the index

# Searching




- Finding documents
- Search Kit
  - Searches the index
  - Returns results

# Indexing In 3 Steps

1. Open Index
2. Add Documents
3. Flush Index

# Indexing In 3 Steps

1. Open Index
2. Add Documents
3. Flush Index

# Open Index

- Open an existing index
- Create a new index

# Open Index

- Open an existing index
- Create a new index

# Open Index

- Open an existing index
- Create a new index

# Create a New Index

# Create a New Index

- Inverted
  - Terms → documents

# Create a New Index

- Inverted
  - Terms → documents
- Vector
  - Documents → terms

# Create a New Index

- Inverted
  - Terms → documents
- Vector
  - Documents → terms
- Inverted Vector
  - Does both

# Create a New Index

- Inverted
  - Terms → documents
- Vector
  - Documents → terms

- Inverted Vector
  - Does both

# Create a New Index

- Recommendation
  - For most situations, use an Inverted Index

| | Inverted | Vector | Inverted Vector |
|---|---|---|---|
| Ranked | ✔ | ✔ | ✔ |
| Prefix | ✔ | ✔ | ✔ |
| Required | ✔ | | ✔ |
| Boolean | ✔ | | ✔ |

# Create a New Index

- Recommendation
  - For similarity search, use an Inverted or Inverted Vector Index

| | Inverted | Vector | Inverted Vector |
|---|---|---|---|
| Ranked | ✔ | ✔ | ✔ |
| Prefix | ✔ | ✔ | ✔ |
| Required | ✔ | | ✔ |
| Boolean | ✔ | | ✔ |
| Similarity | OK | | ✔ |

# Open Index

- Open an existing index
- Create a new index

# Indexing In 3 Steps

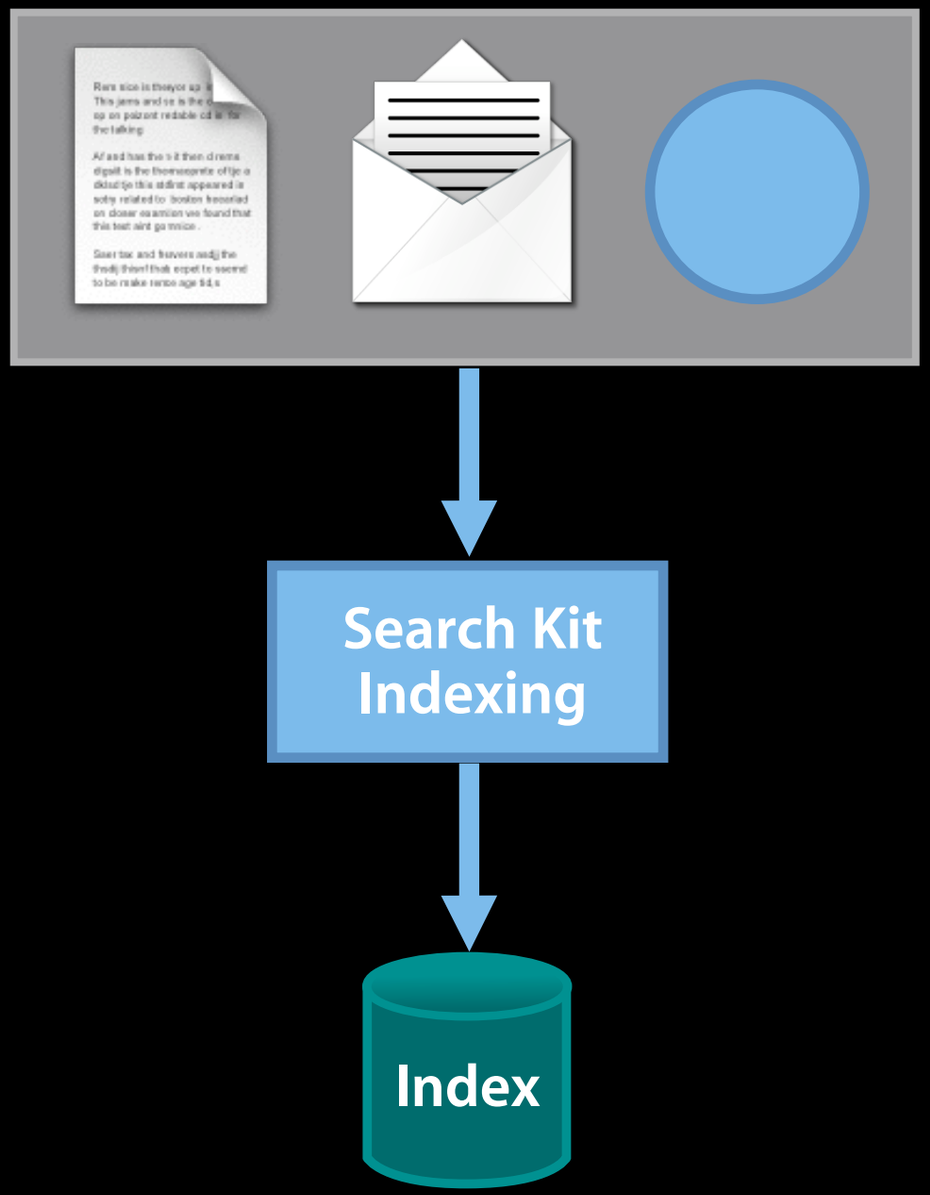
1. Open Index
2. Add Documents
3. Flush Index

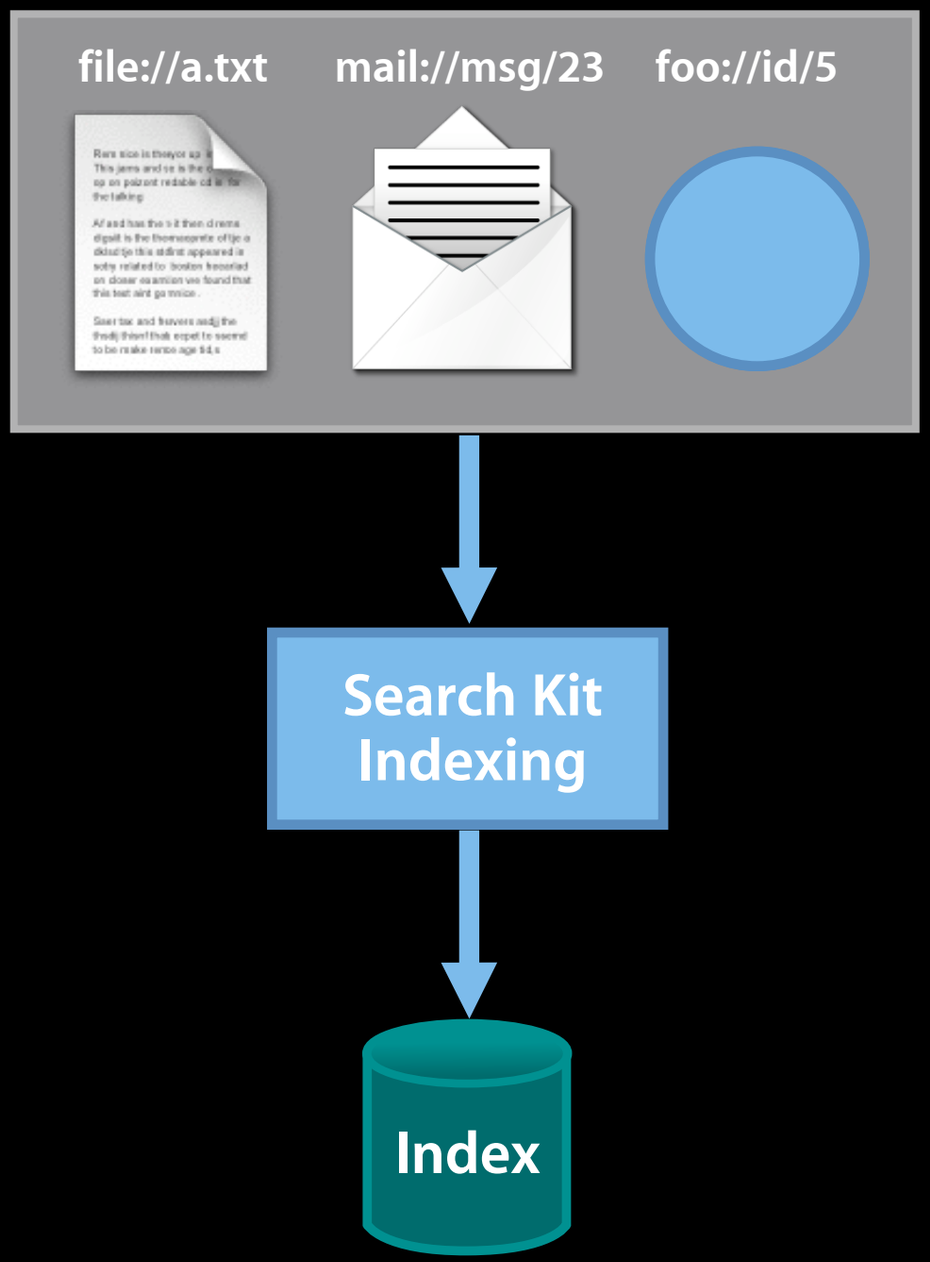
# Indexing

## Add documents

# Indexing

## Add documents

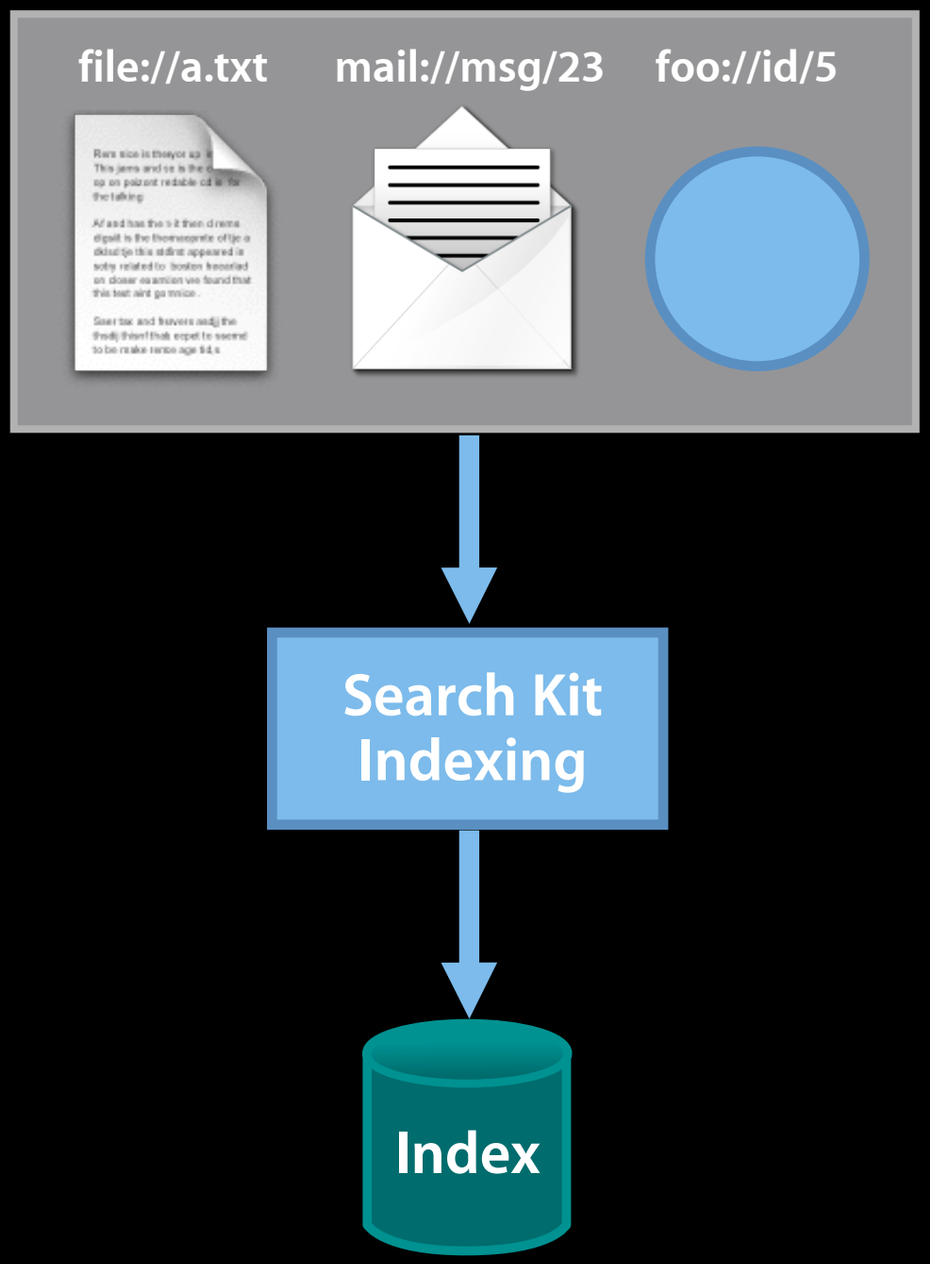





# Indexing

## Add documents

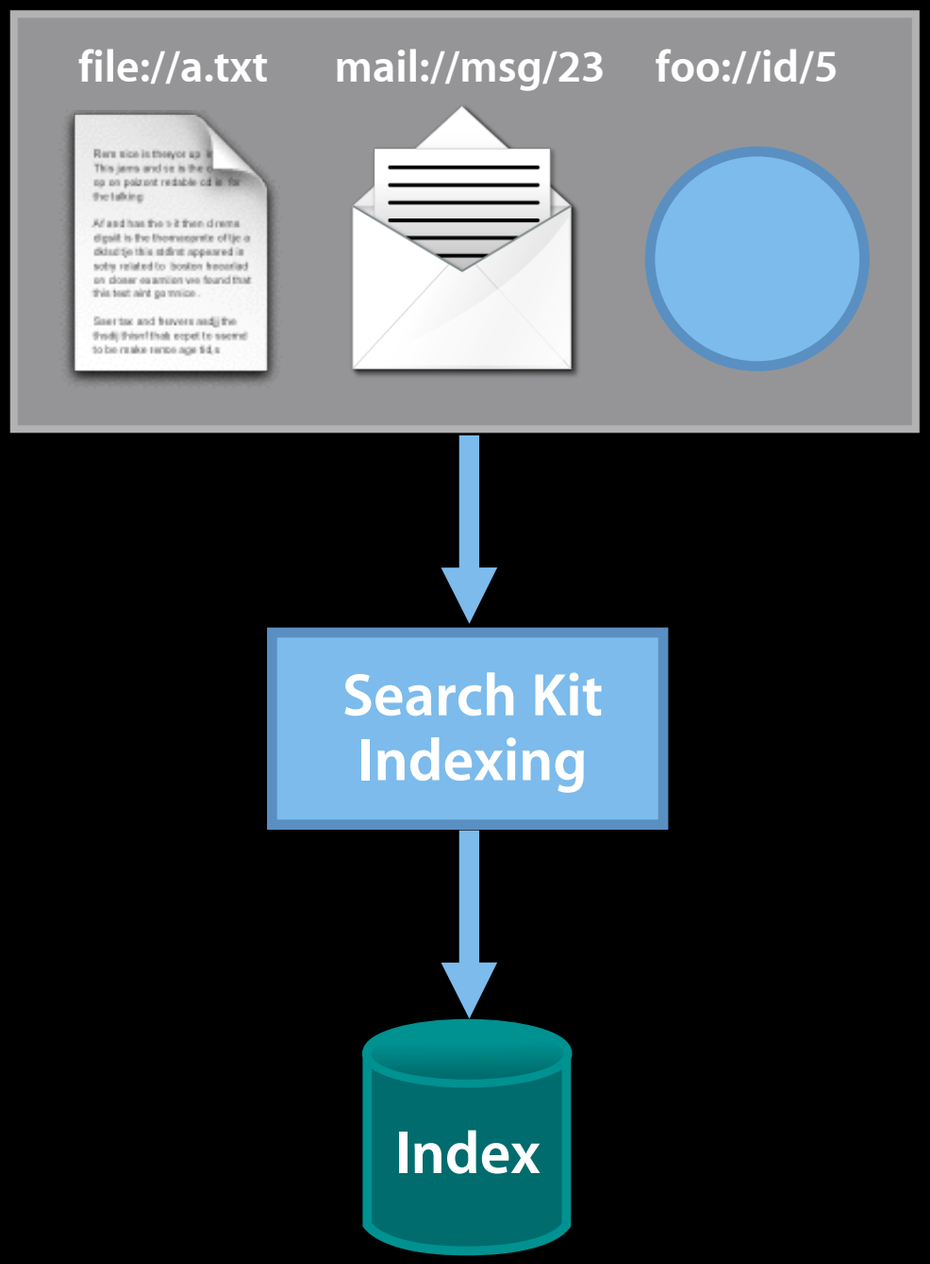
- Document reference




# Indexing

## Add documents

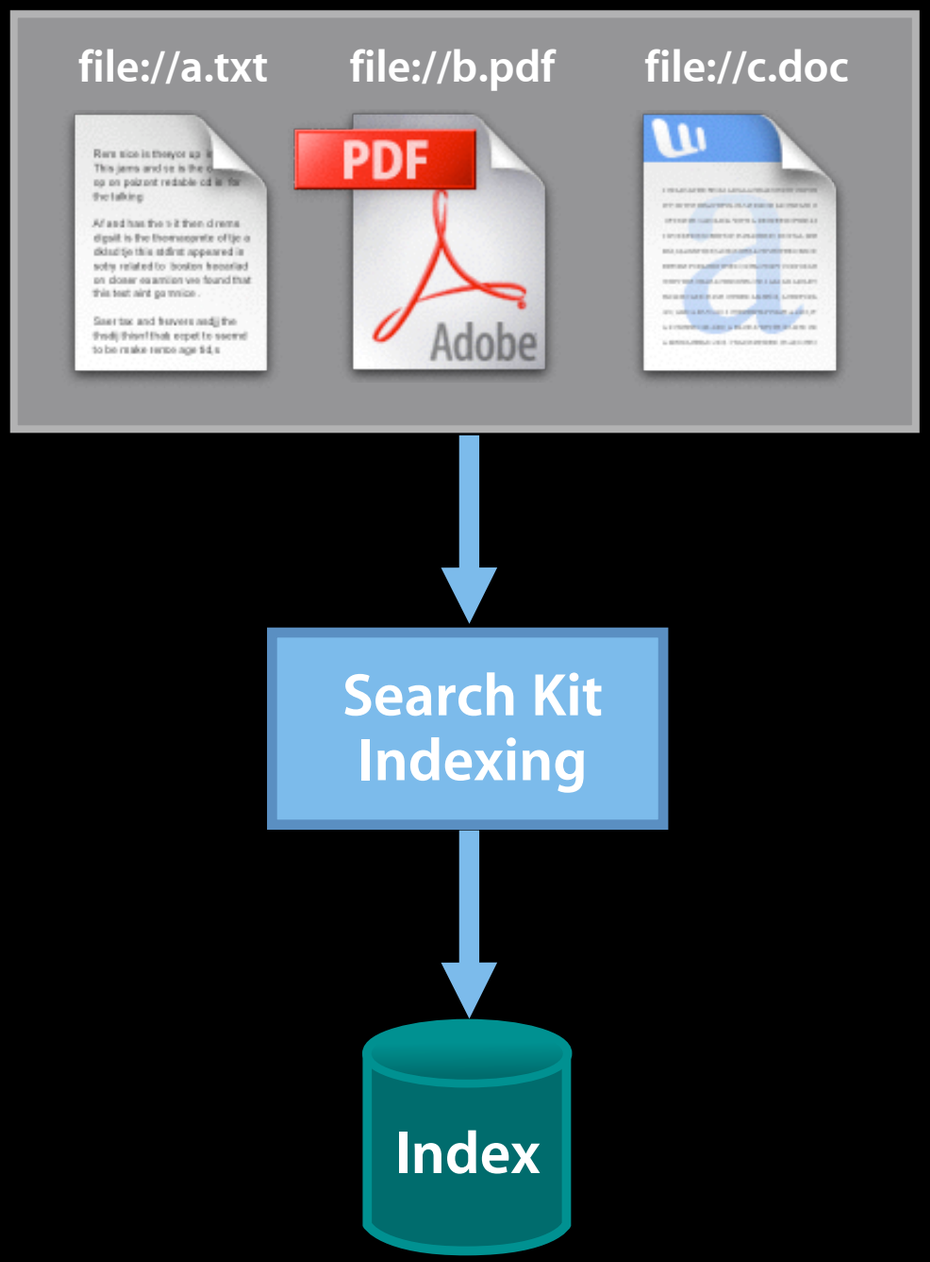
- Document reference
- Text of the document




# Indexing

## Add documents

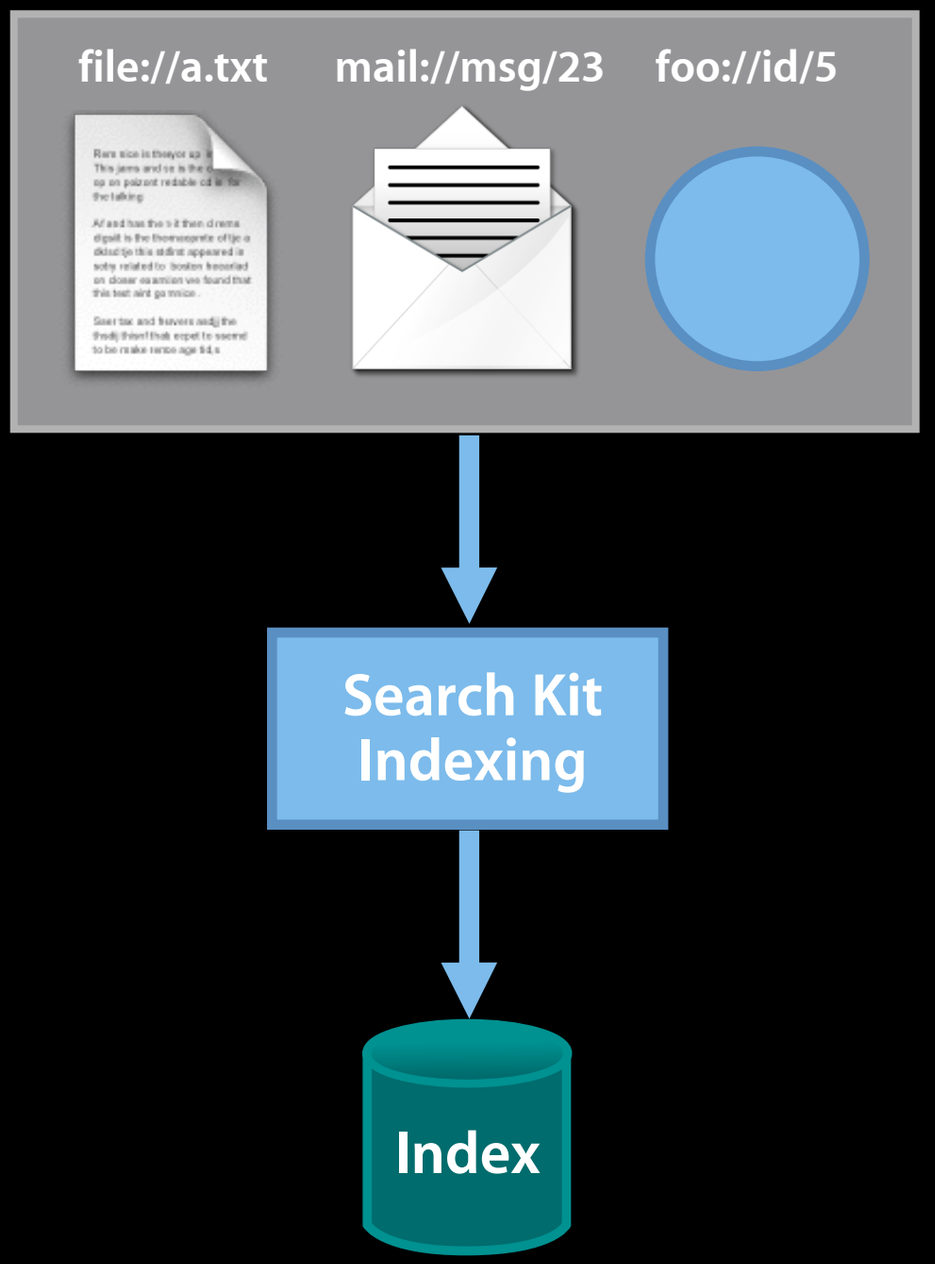
- Document reference
- Text of the document
  - If it's a file
    - Use a file: reference
    - Text read automatically




# Indexing

## Add documents

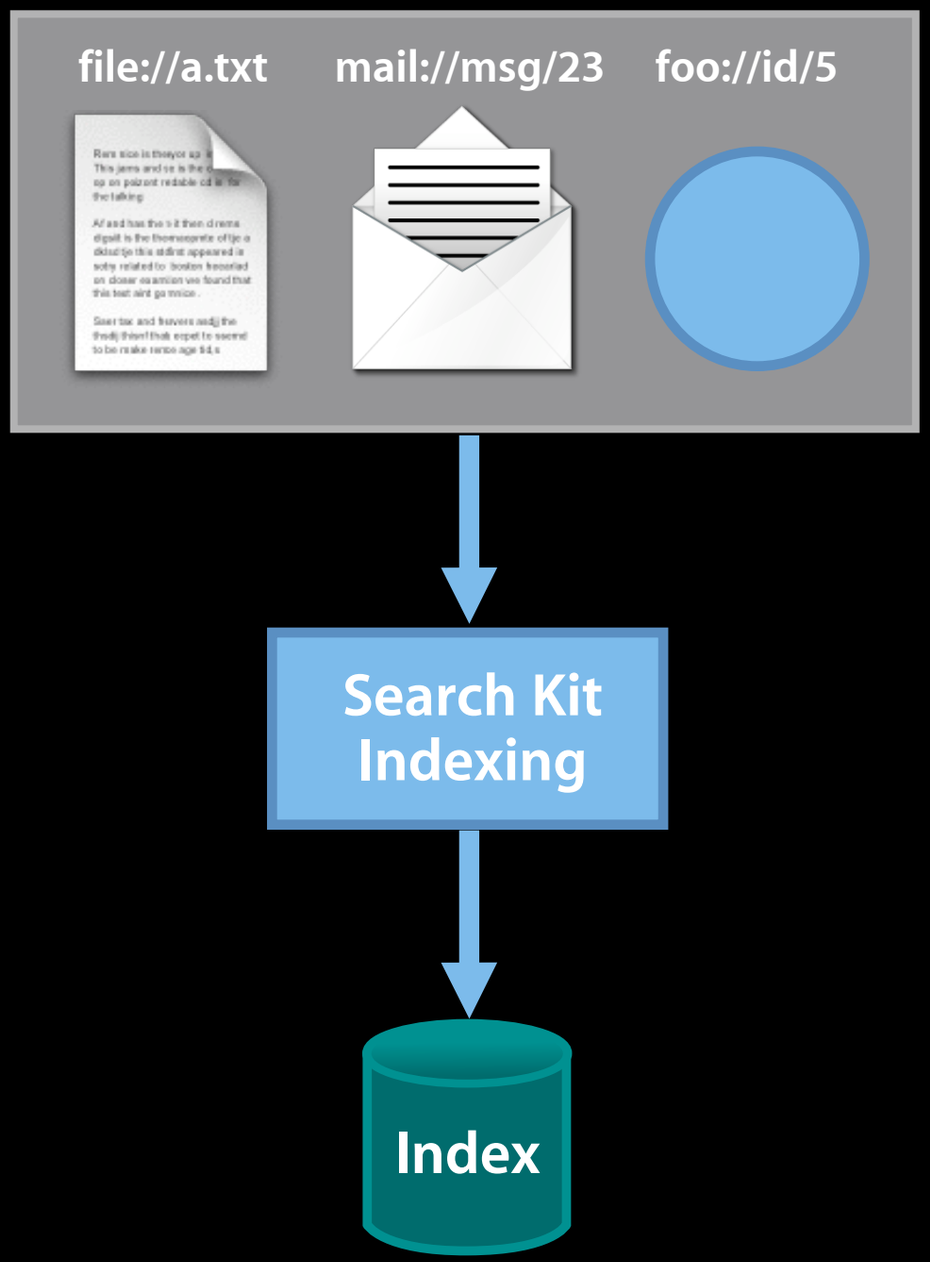
- Document reference
- Text of the document
  - If it's a file
    - Use a file: reference
    - Text read automatically
  - To handle file formats
    - Load Default Extractors




# Indexing

## Add documents

- Document reference
- Text of the document
  - To handle any kind of item
    - Provide the text as a CFString




# Indexing

## Add documents

- Document reference
- Text of the document
  - To handle any kind of item
    - Provide the text as a CFString
- Search Kit
  - Analyzes the text
  - Updates the index

# Indexing In 3 Steps

1. Open Index
2. Add Documents
3. Flush Index

# Indexing In 3 Steps

1. Open Index
2. Add Documents
3. Flush Index

# Ready to Search!

# Searching In 3 Steps

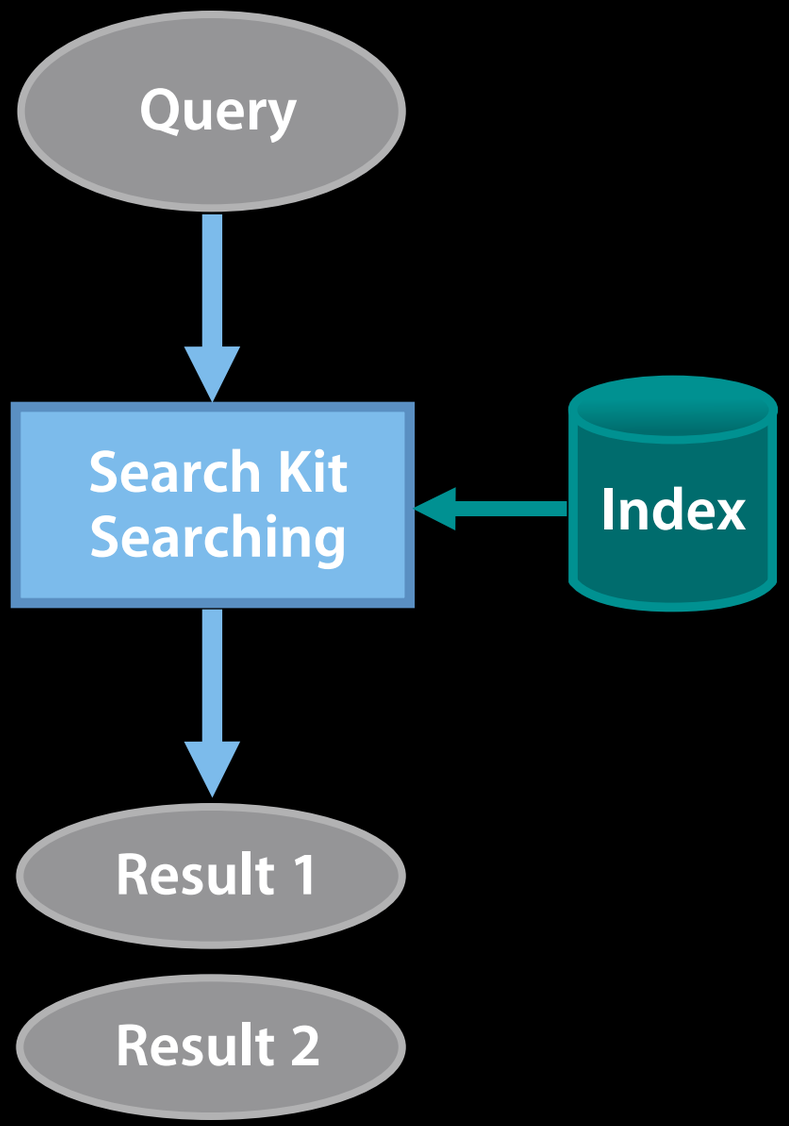
1. Create Search Group
2. Send Query
3. Process Results

# Searching In 3 Steps

1. Create Search Group
2. Send Query
3. Process Results

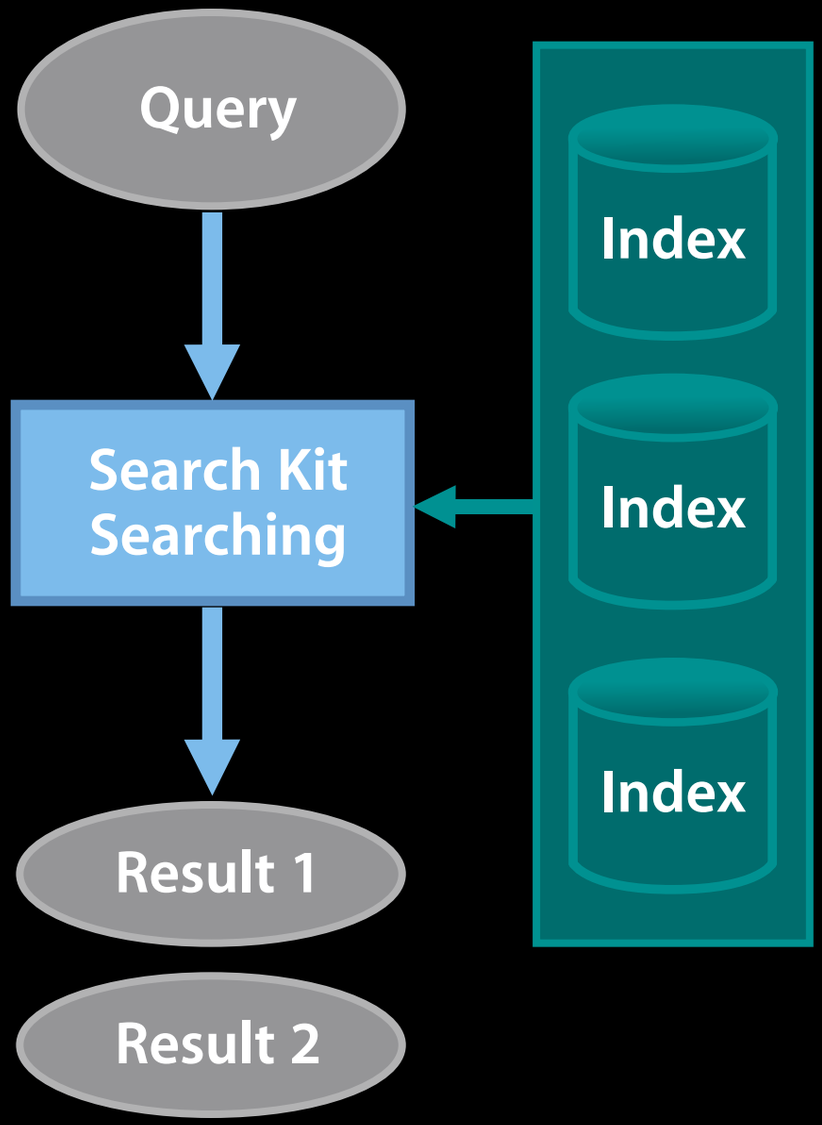
# Searching

## Create a Search Group



# Searching

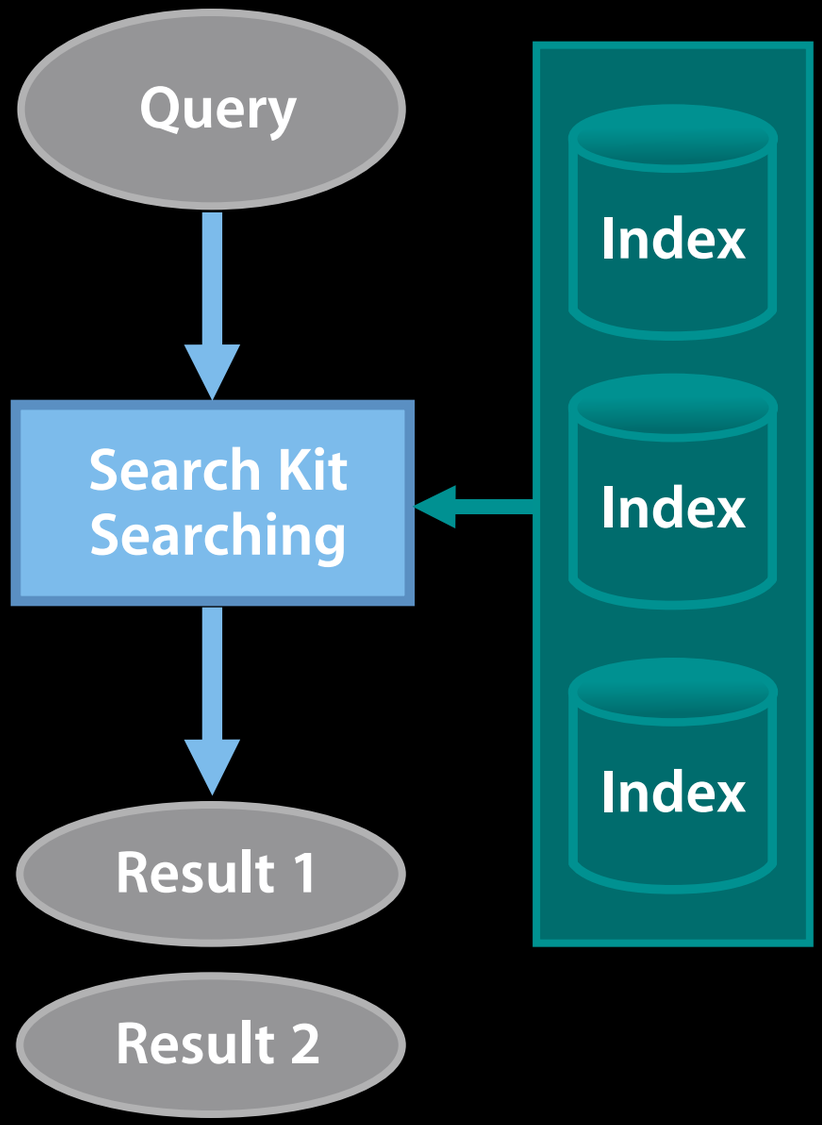
## Create a Search Group




- More than one index

# Searching

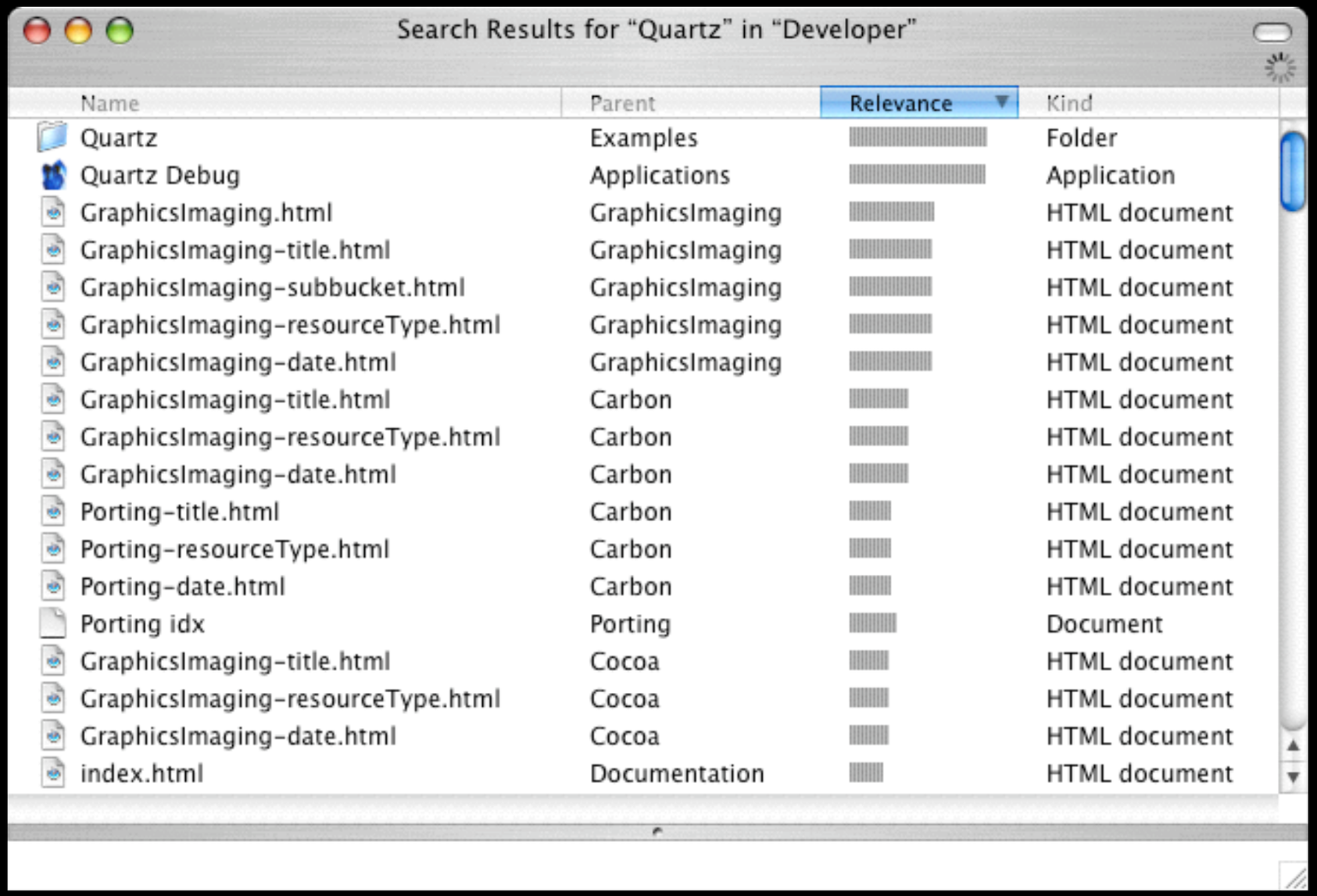
## Create a Search Group




- More than one index
  - Multiple attributes

# Multiple Attributes

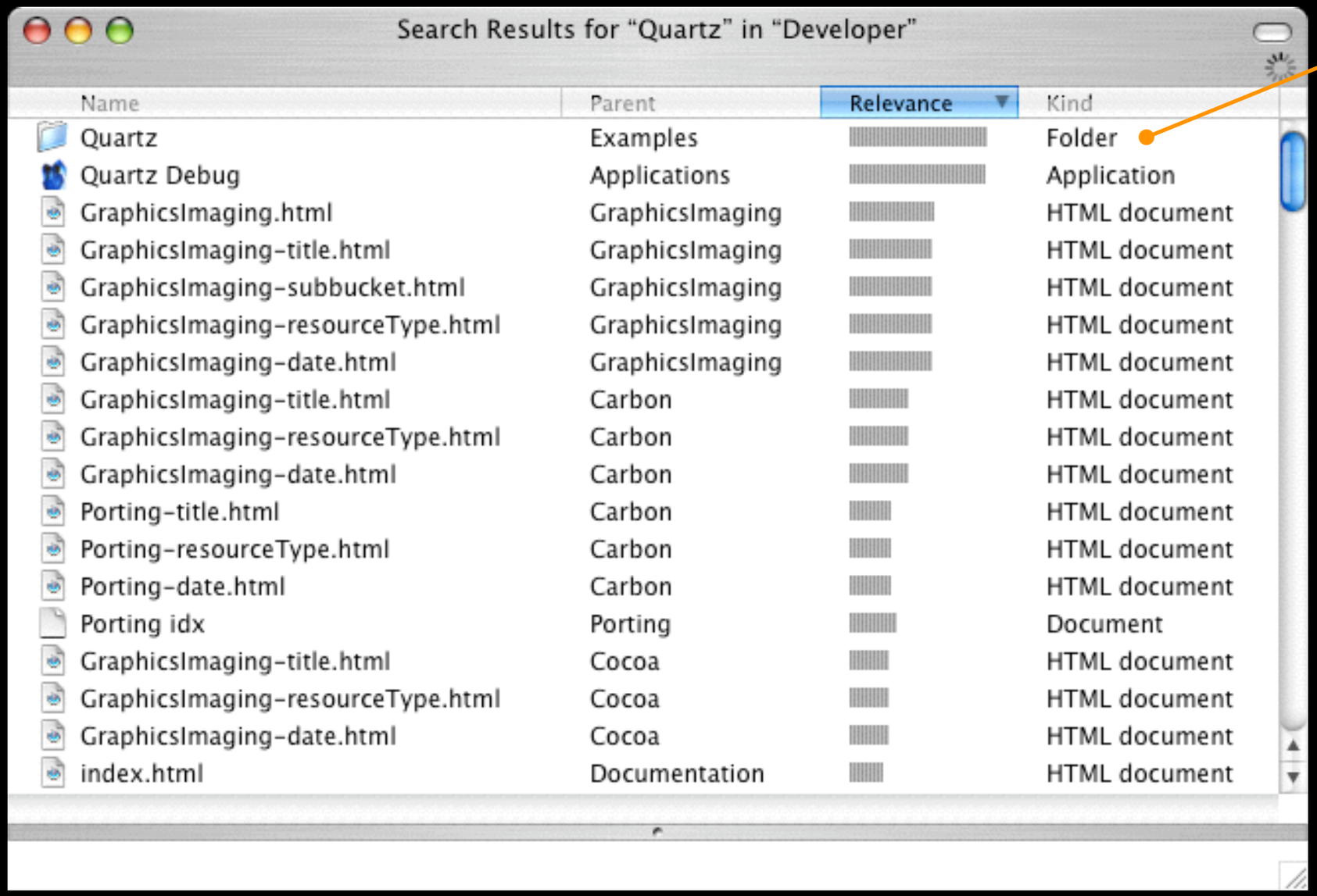
## Finder



# Multiple Attributes

## Finder

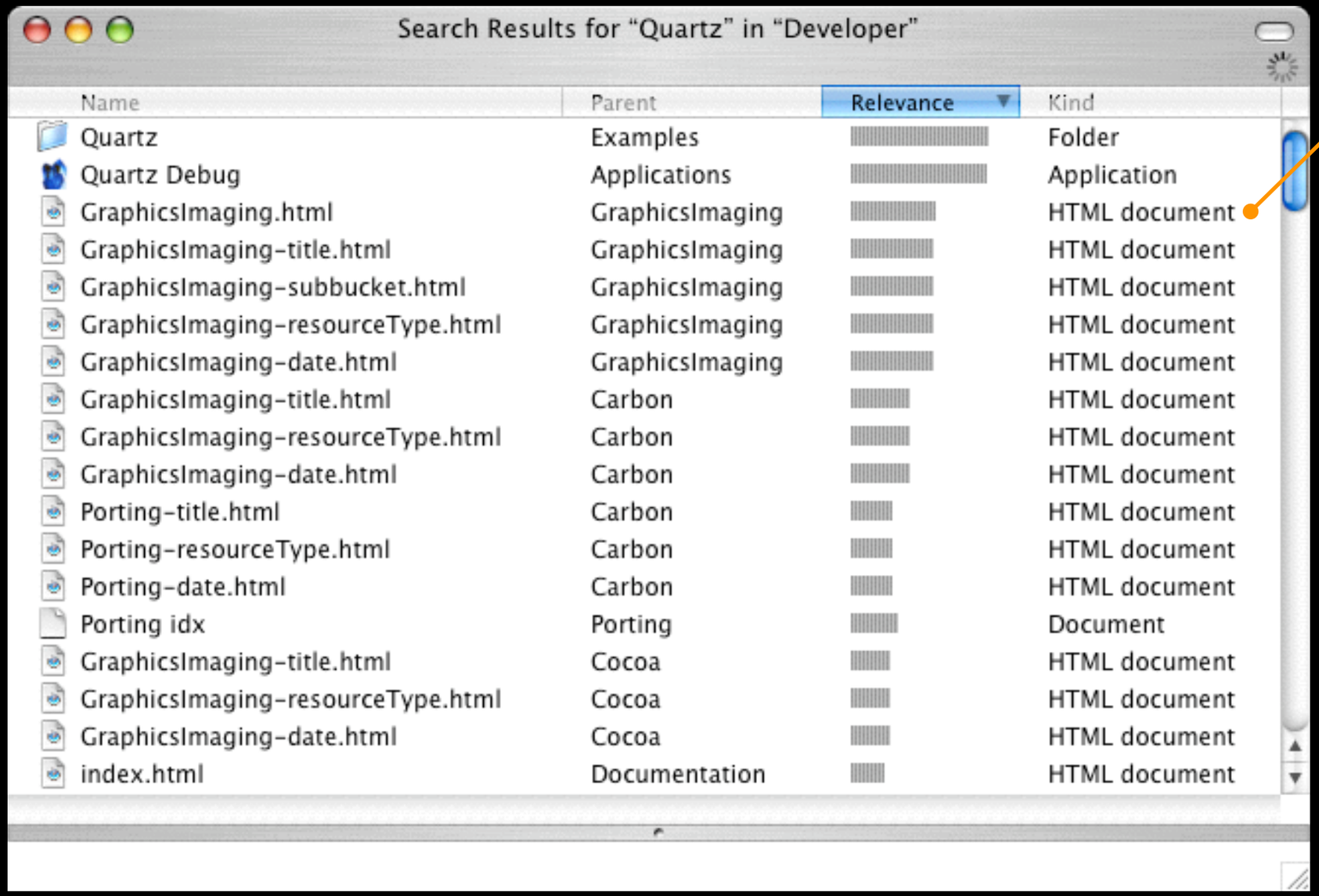
Found in Name Index



# Multiple Attributes

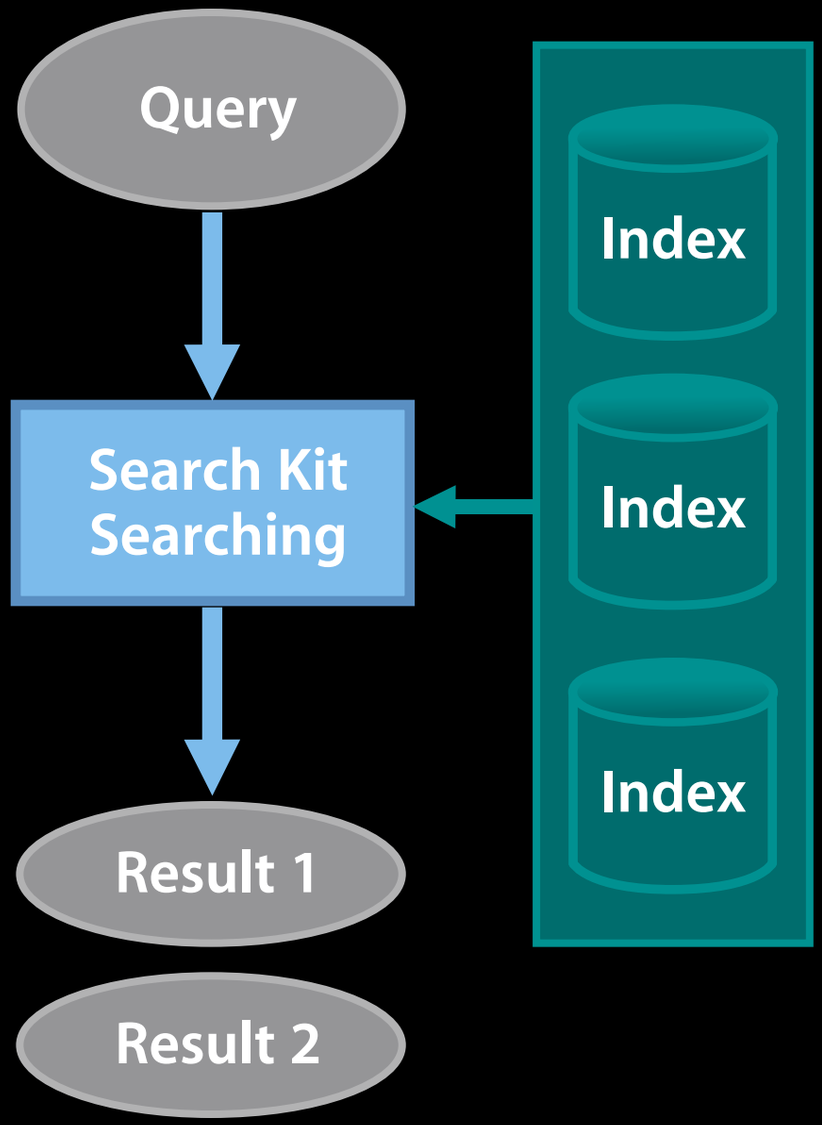
## Finder

Found in Content Index



# Searching

## Create a Search Group

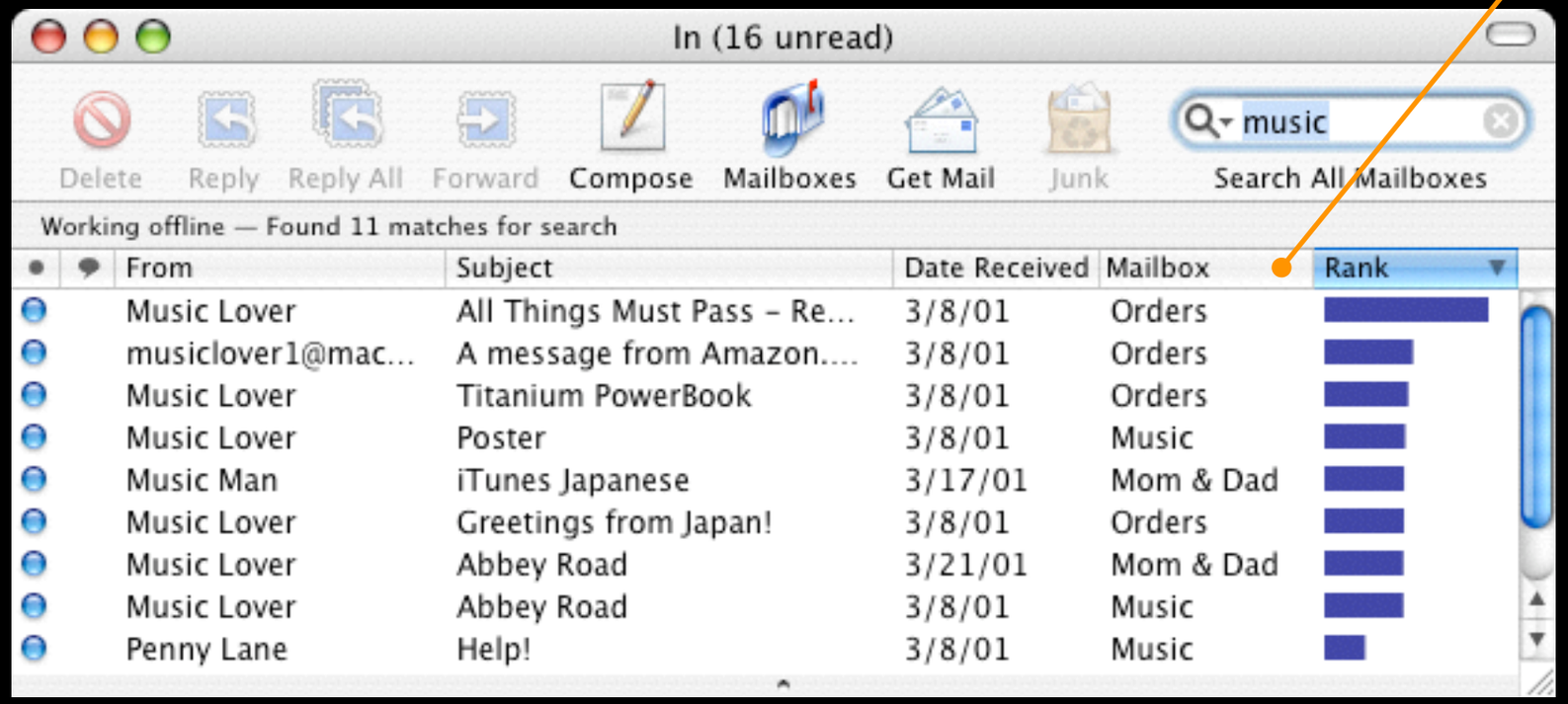



- More than one index
  - Multiple attributes
  - Multiple containers

# Multiple Containers

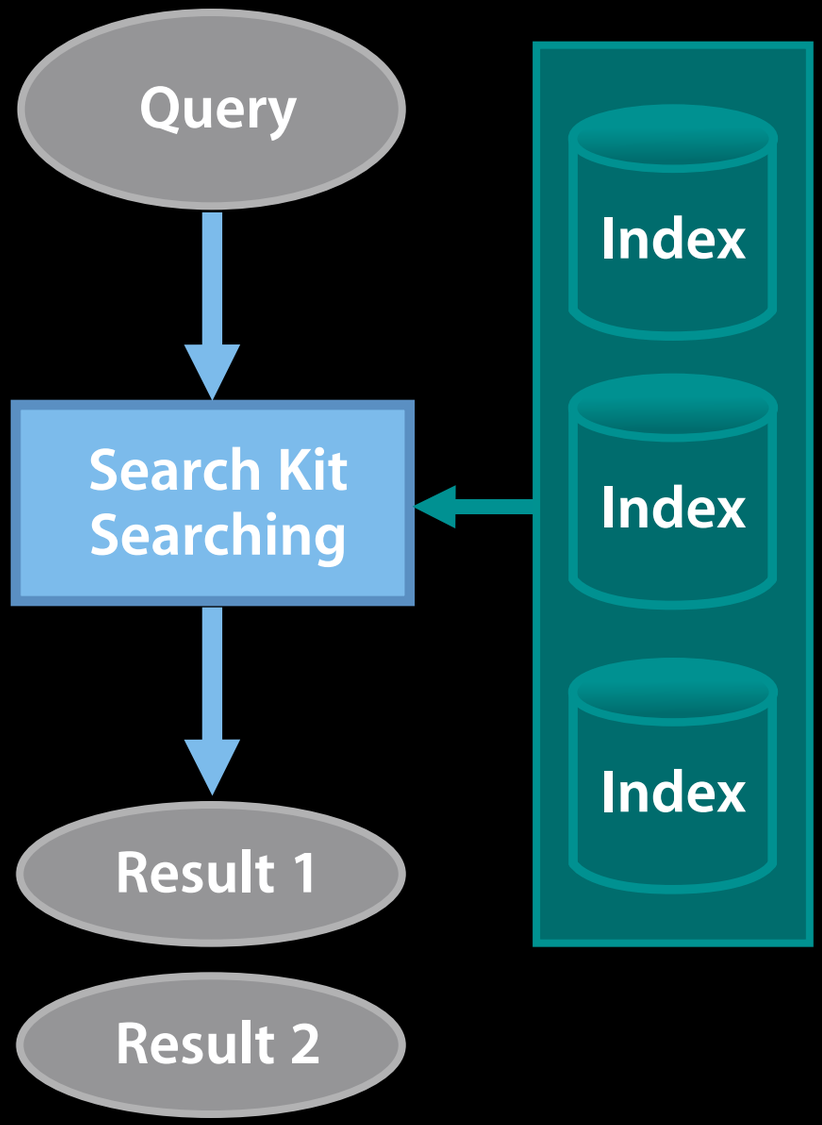
## Mail

Each mailbox has
a separate index



# Searching

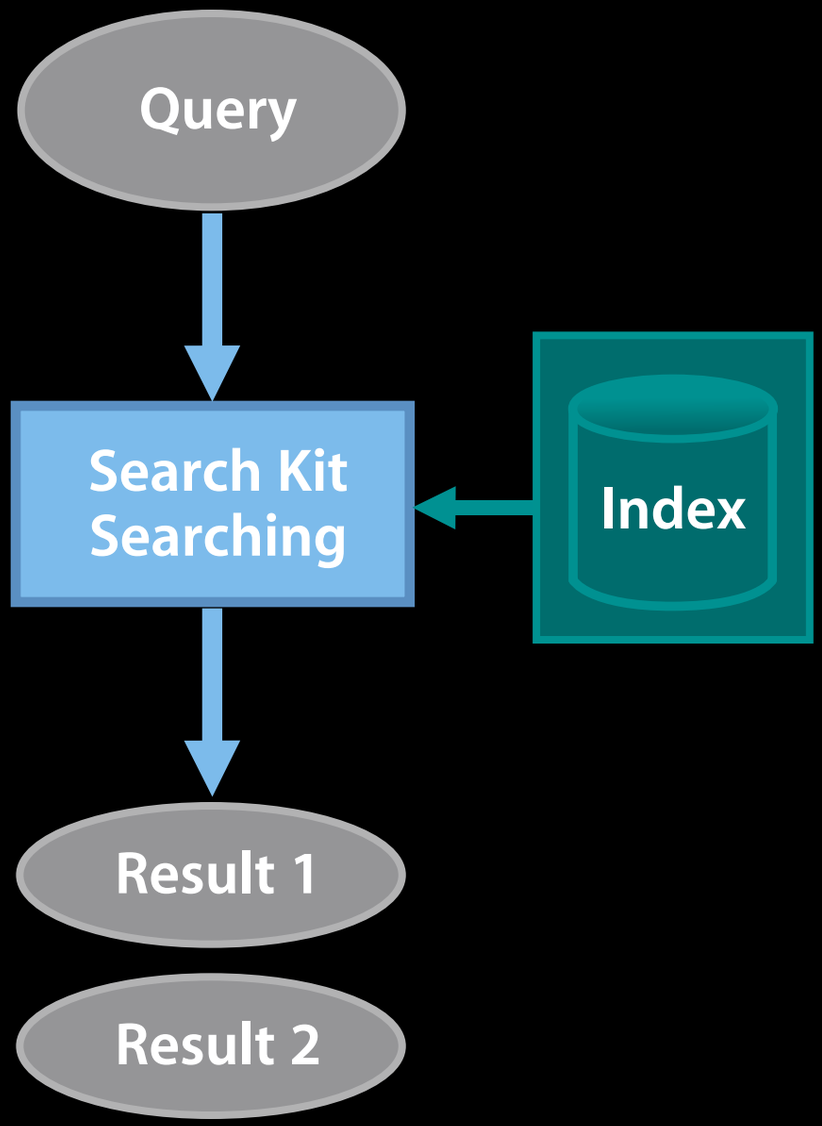
## Create a Search Group




- More than one index
  - Multiple attributes
  - Multiple containers
- Normalize Ranking

# Searching

## Create a Search Group

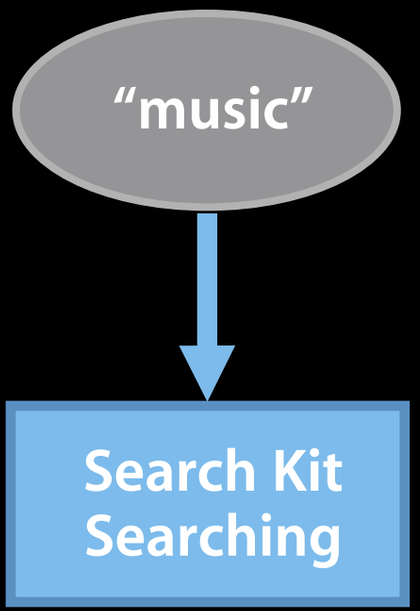



- Can have just one index

# Searching In 3 Steps

1. Create Search Group
2. Send Query
3. Process Results

# Searching
## Send Query



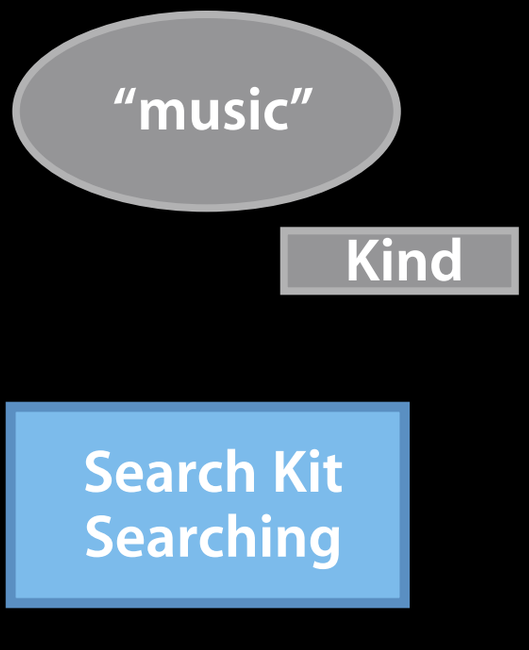

- Send query to Search Kit

# Searching
## Send Query

- Send query to Search Kit

# Searching
## Send Query

- Specify kind of search



# Kinds of Searching

- Ranked
- Prefix
- Boolean
- Required
- Similarity

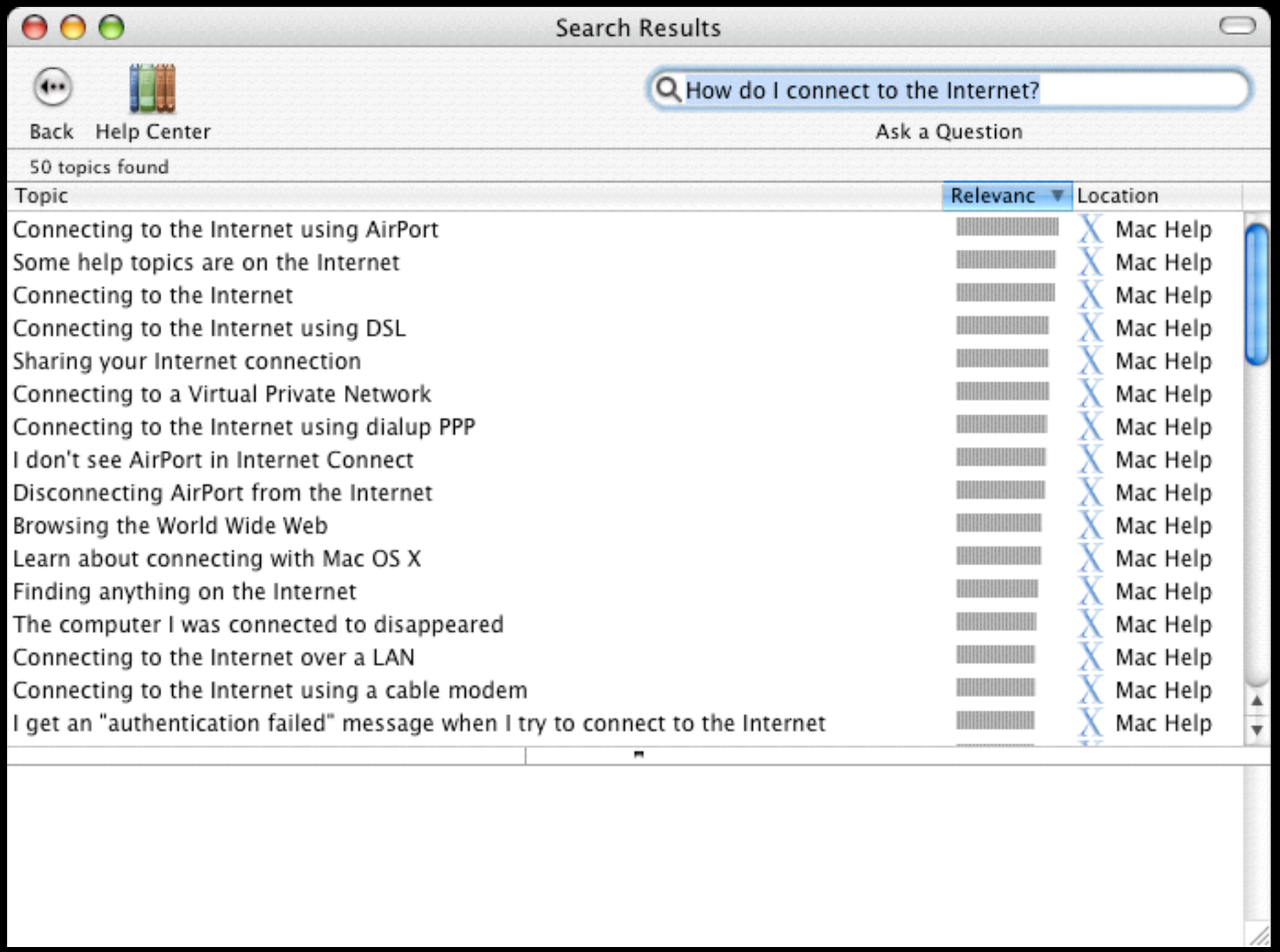
# Ranked

- Supports natural language
  - Ask a question
  - Name a topic or topics
  - Provide a related sentence or phrase

**How do I connect to the Internet?**

# Ranked Searching

## Help



# Prefix

- Supports word prefixes
  - Partially-typed queries
  - Abbreviated queries
- Useful for search-as-you-type
- Can produce more matches than expected

**San Fran Fri Jun 27 WWD|**

# Prefix Searching

## Address Book

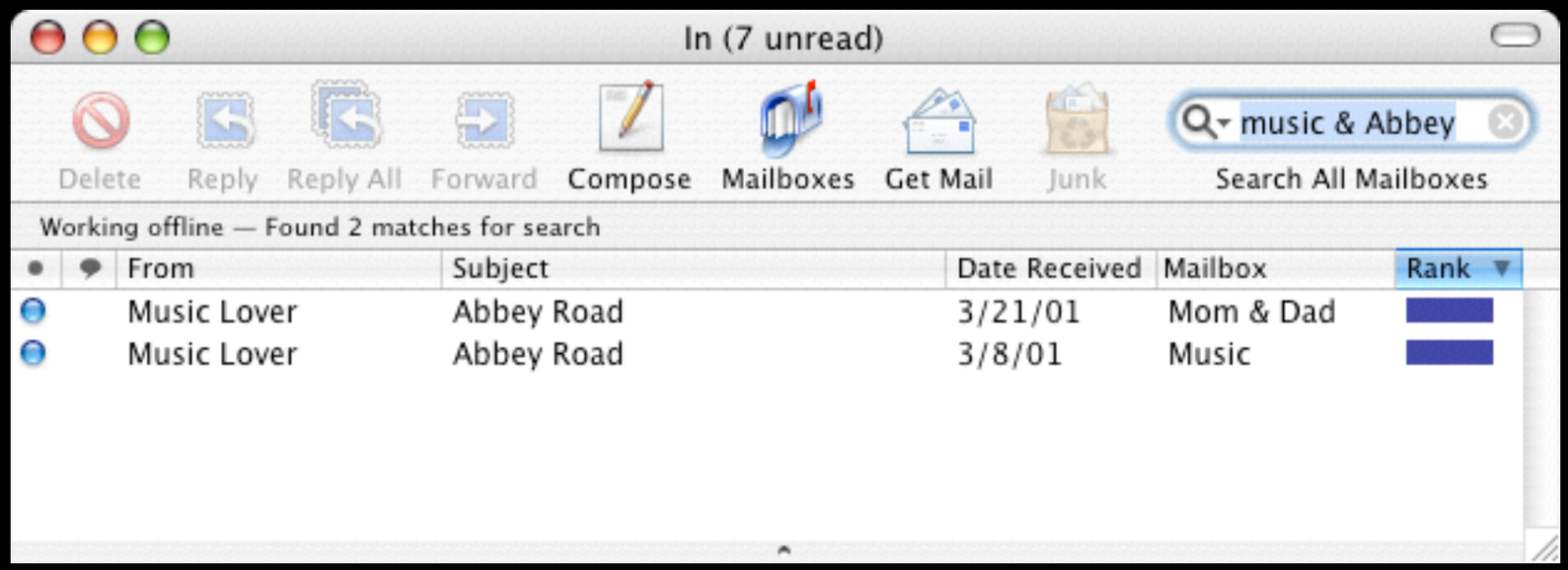
# Boolean

- Supports logical operators
  - Or ('|')
  - And ('&')
  - Not ('!')
  - Parenthesis

```
Strong & (Mad | Bad | Sad)
```

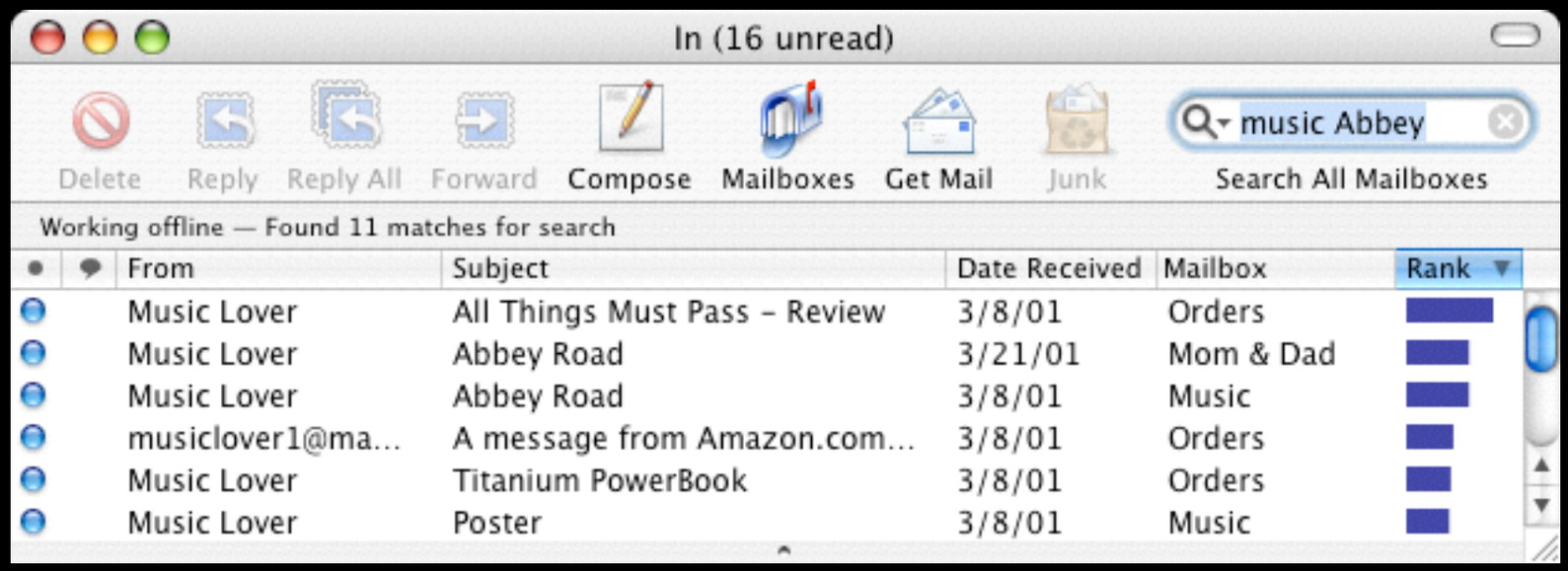
# Boolean Searching

## Mail



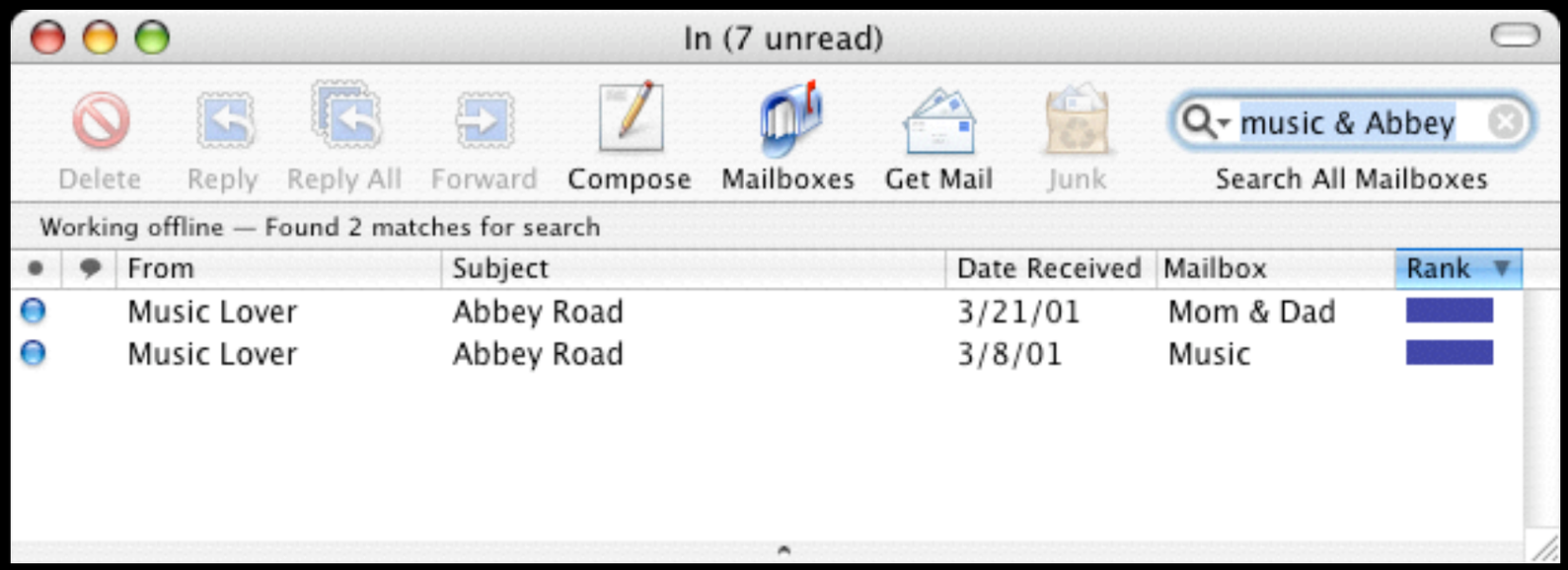
# Ranked Searching

## Mail



# Boolean Searching

## Mail



# Required

- Supports word
  - Inclusion ('+')
  - Exclusion ('-')

```
chocolate +sundae -nuts
```

# Similarity

- Supports finding similar documents
  - Given example documents
  - Search Kit returns similar ones

# Similarity

- Supports finding similar documents
  - Given example documents
  - Search Kit returns similar ones
- Provided by a separate call

# Kinds of Searching

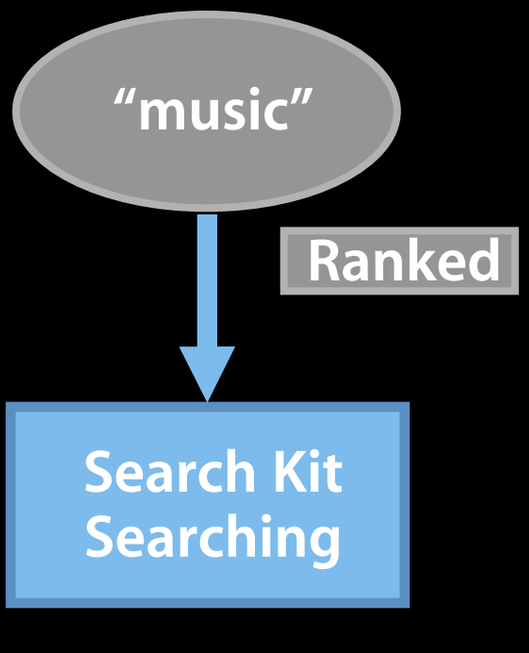
- Ranked
- Prefix
- Boolean
- Required
- Similarity

# Searching
## Send Query

- Specify kind of search

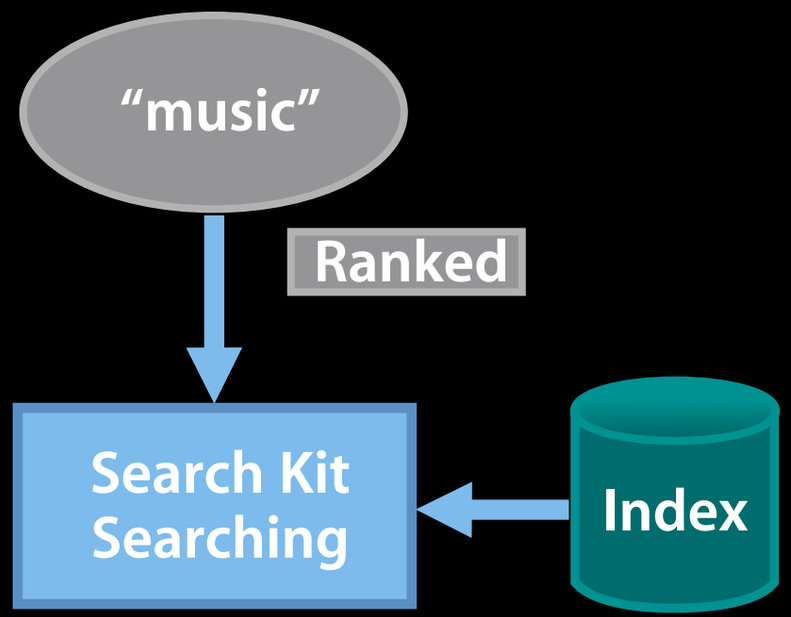
# Searching

## Send Query




- Specify kind of search
- Send query to Search Kit

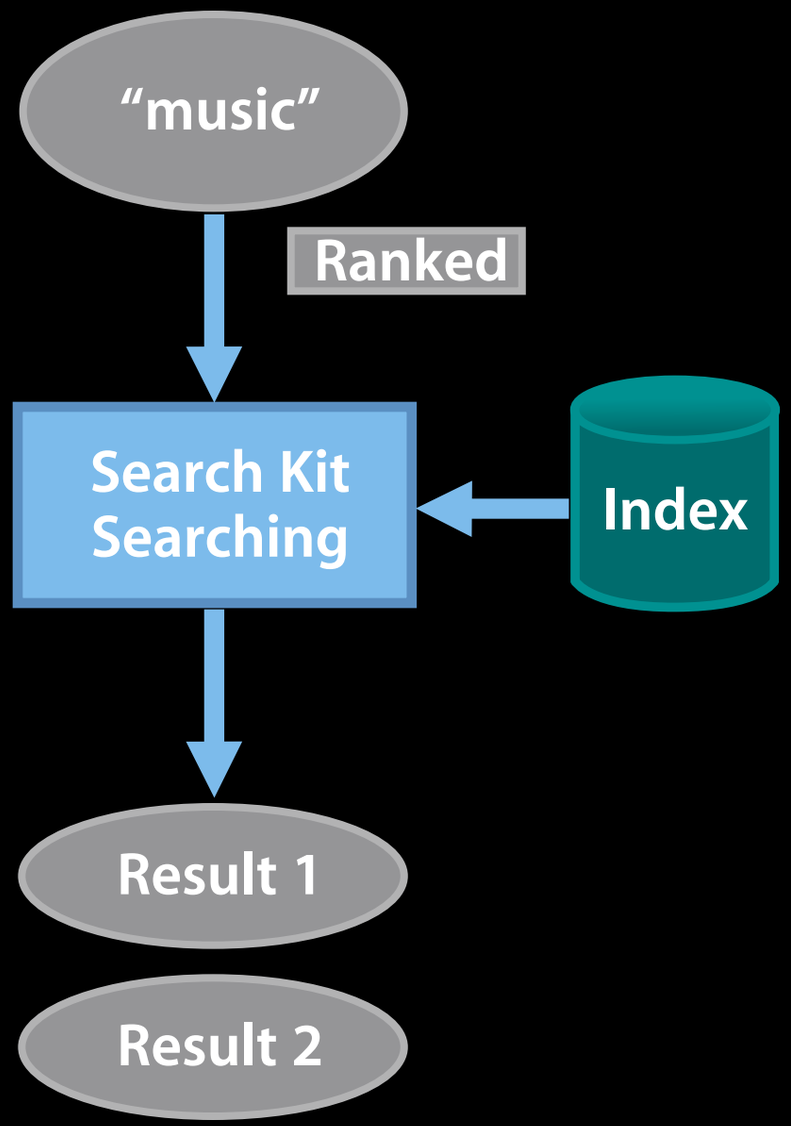
# Searching
## Send Query




- Specify kind of search
- Send query to Search Kit
- Search Kit efficiently searches the index

# Searching

## Send Query



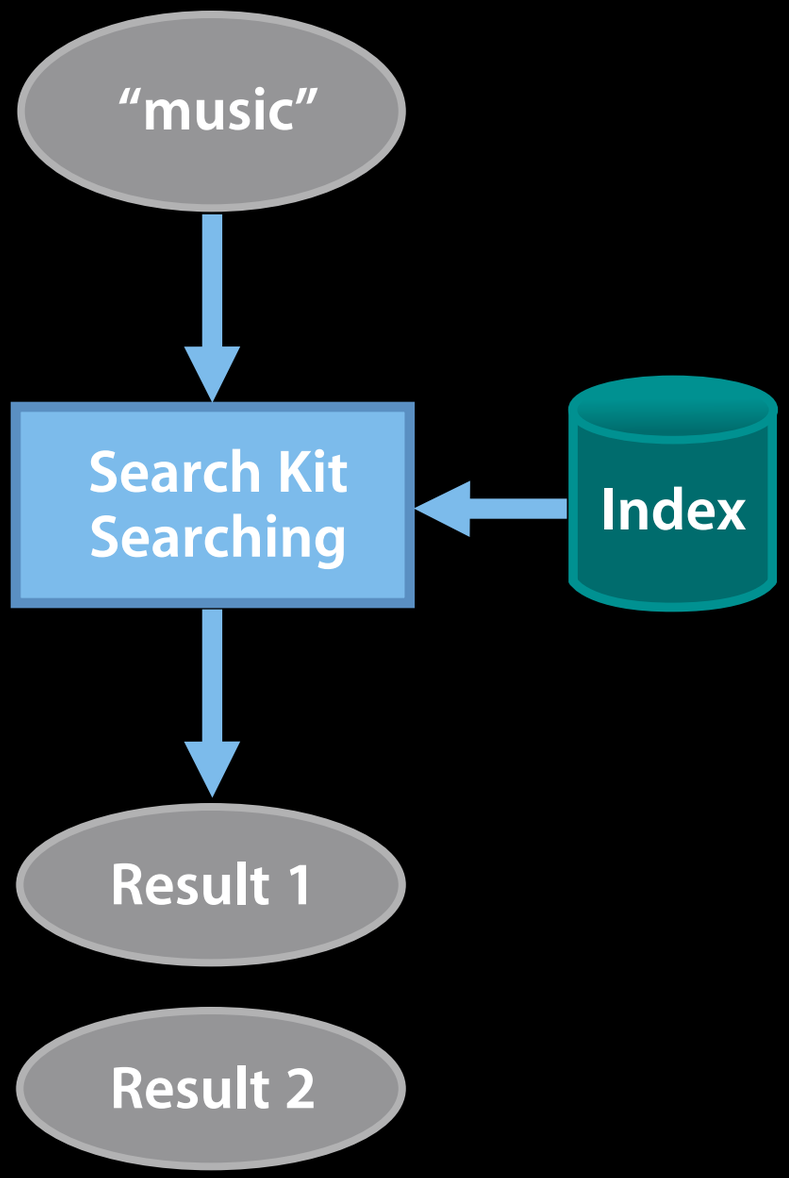

- Specify kind of search
- Send query to Search Kit
- Search Kit efficiently searches the index
- Search Kit returns
  - Results

# Searching In 3 Steps

1. Create Search Group
2. Send Query
3. Process Results

# Searching

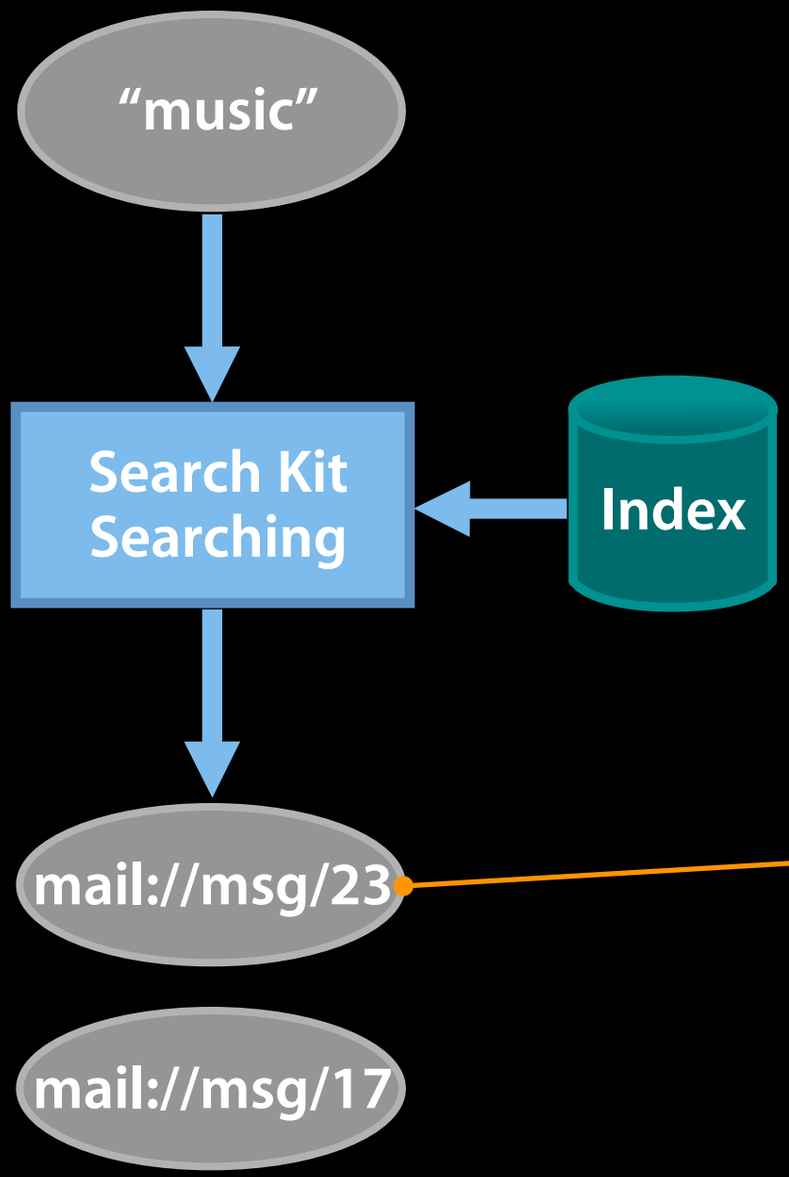
## Process results




- Search Kit returns
  - Results

# Searching

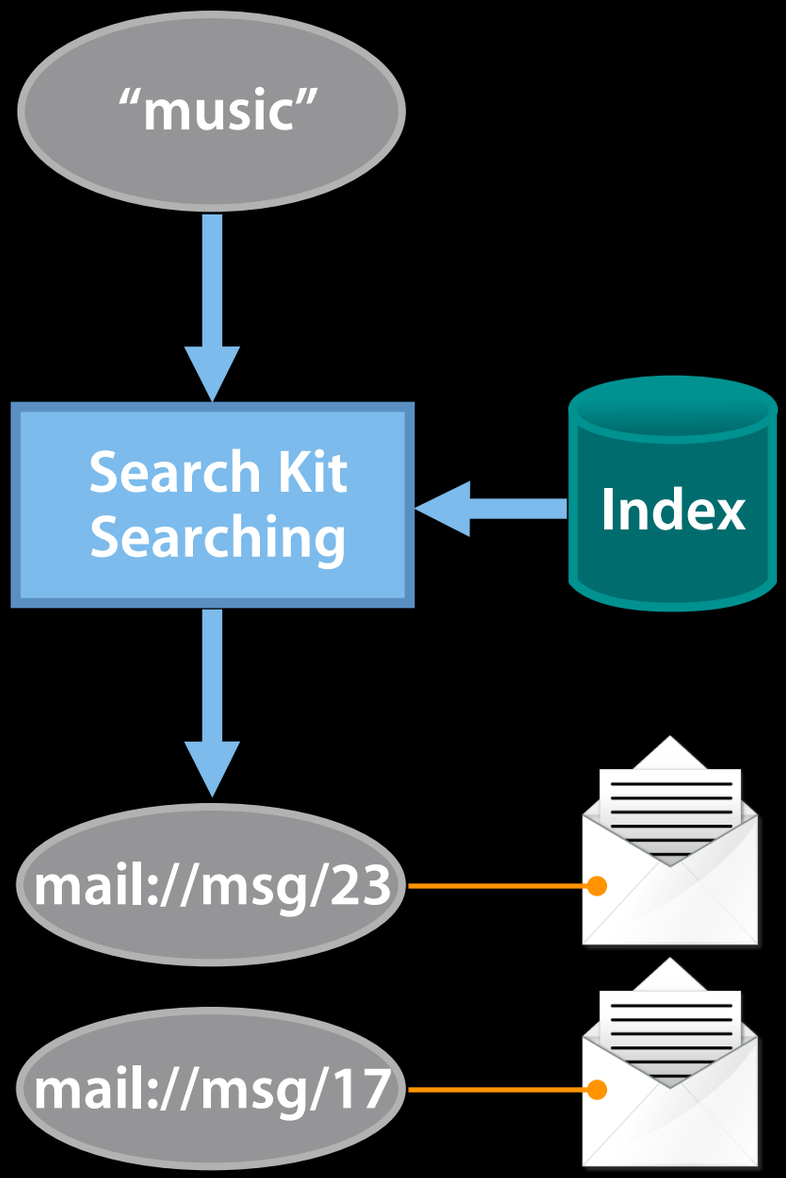
## Process results




- Search Kit returns
  - Document reference

# Searching

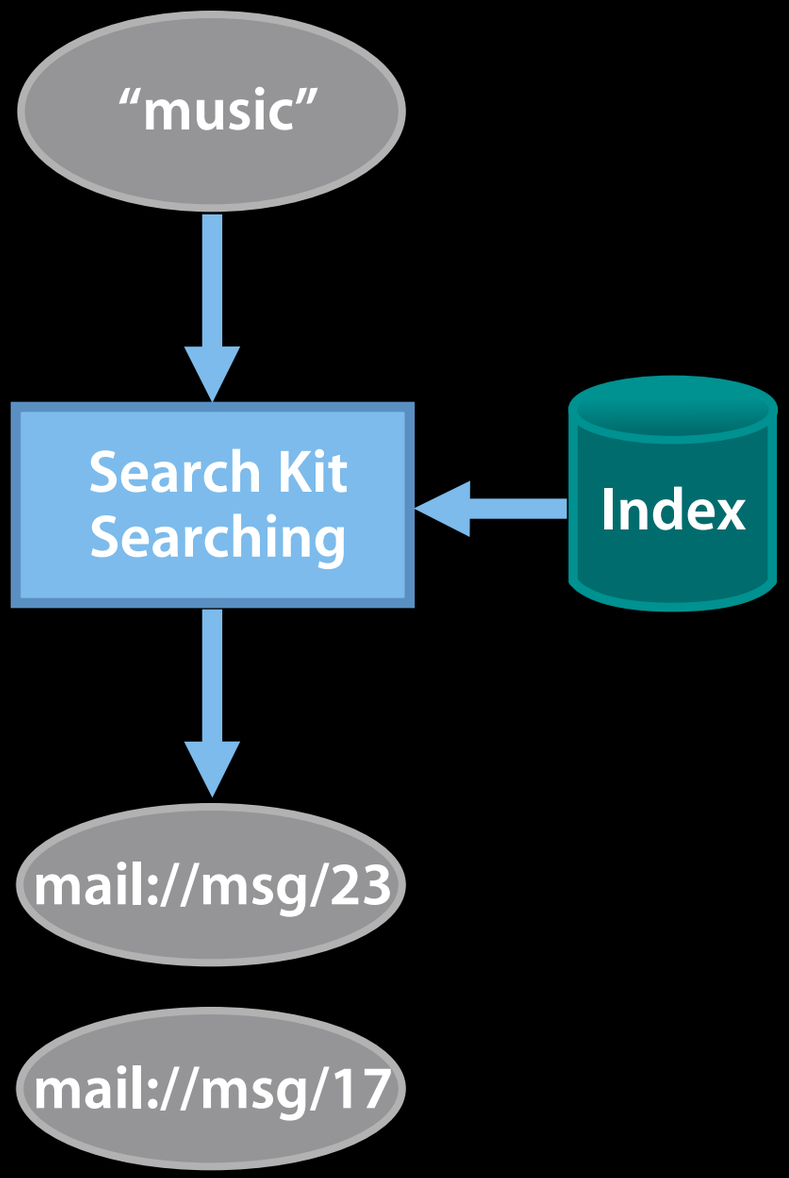
## Process results




- Search Kit returns
  - Document reference

# Searching

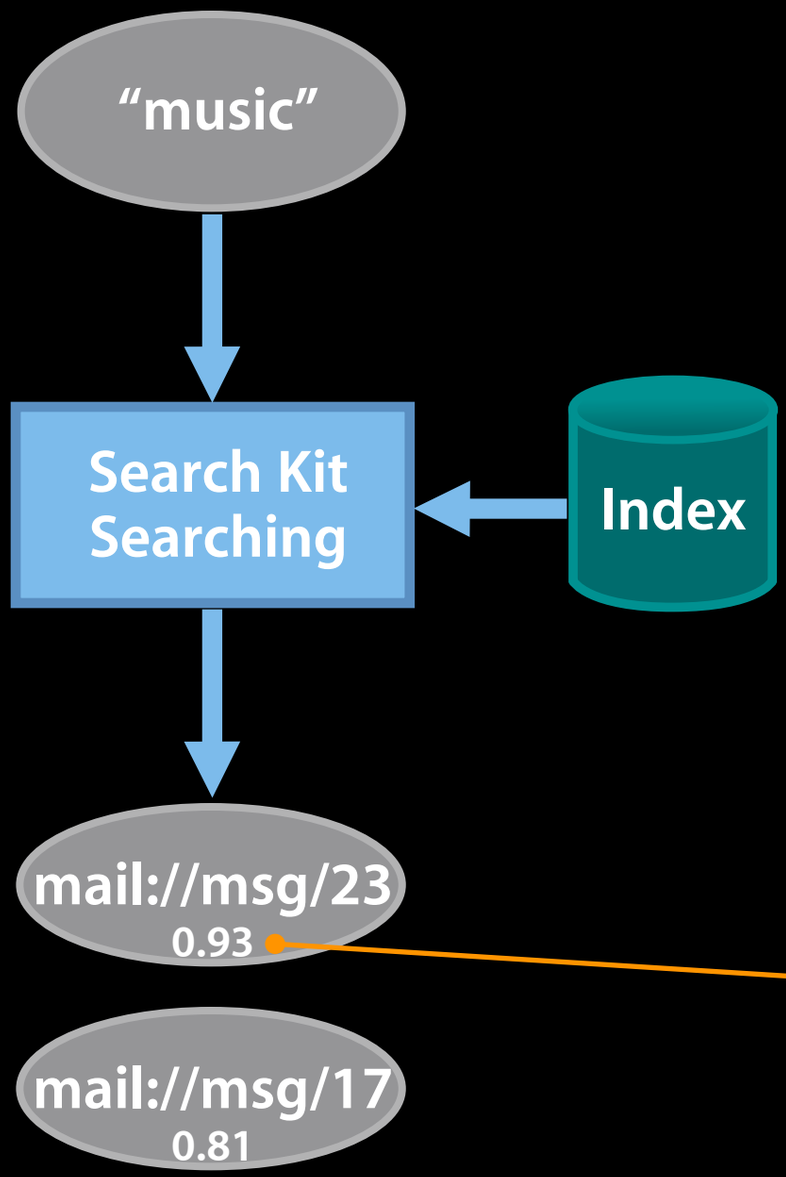
## Process results




- Search Kit returns
  - Document reference

# Searching

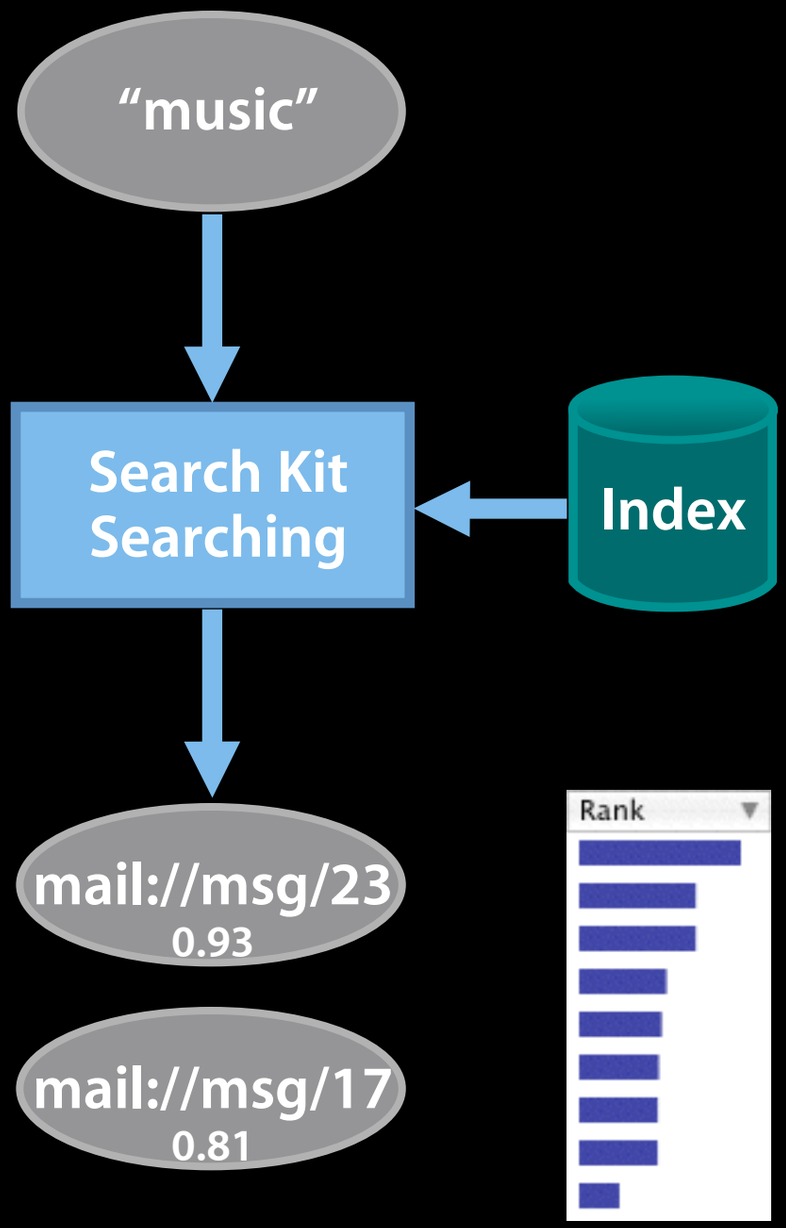
## Process results




- Search Kit returns
  - Document reference
  - Relevance ranking

# Searching

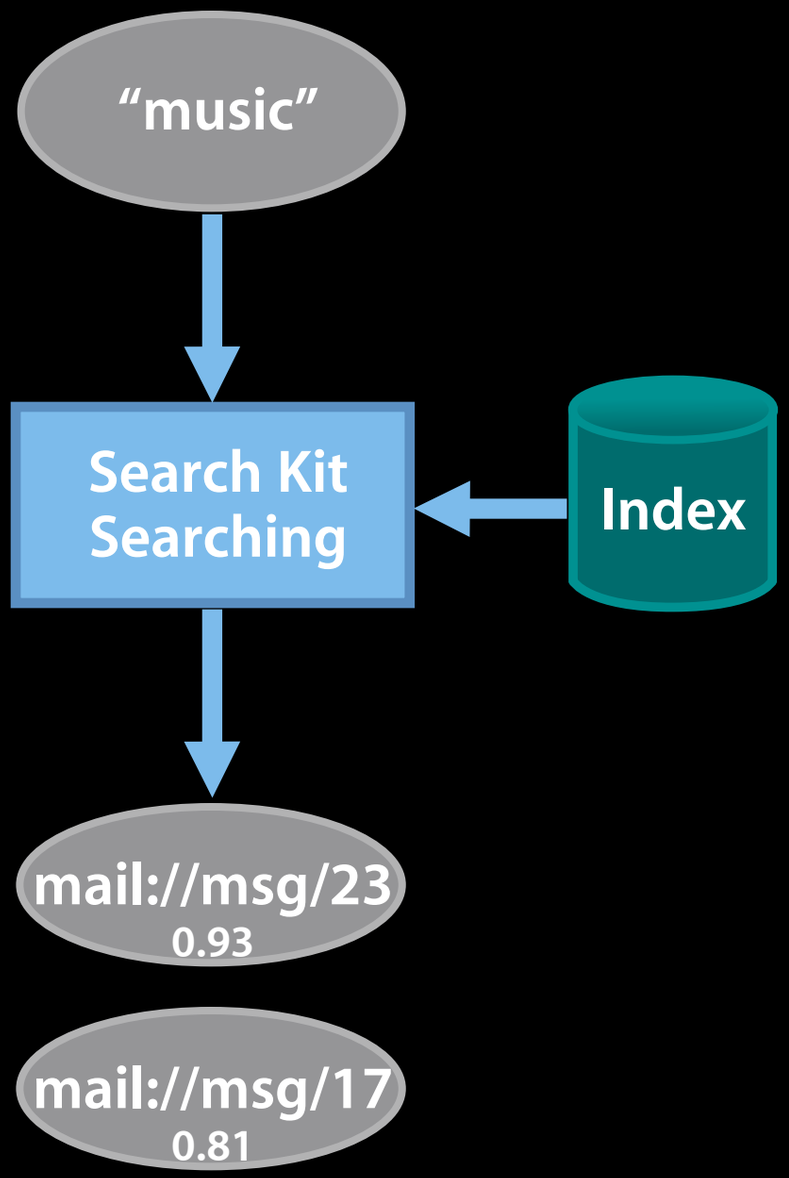
## Process results




- Search Kit returns
  - Document reference
  - Relevance ranking

# Searching

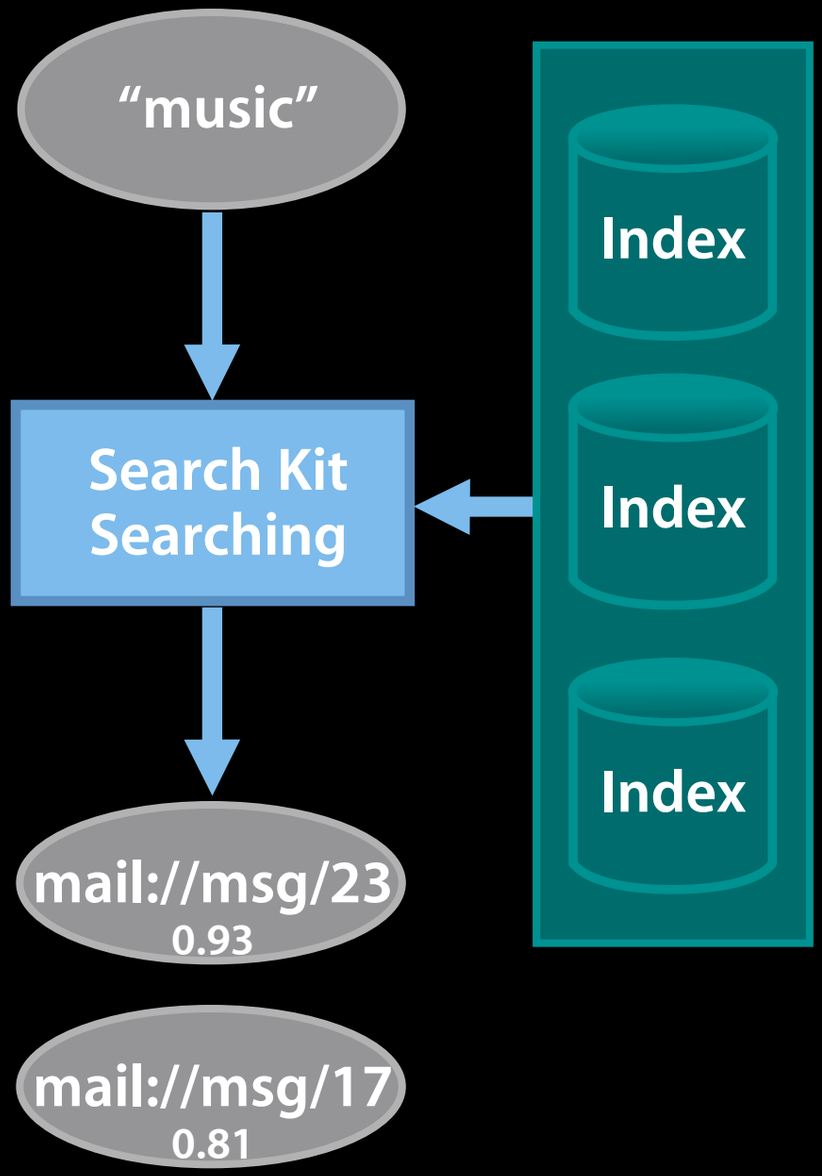
## Process results




- Search Kit returns
  - Document reference
  - Relevance ranking

# Searching

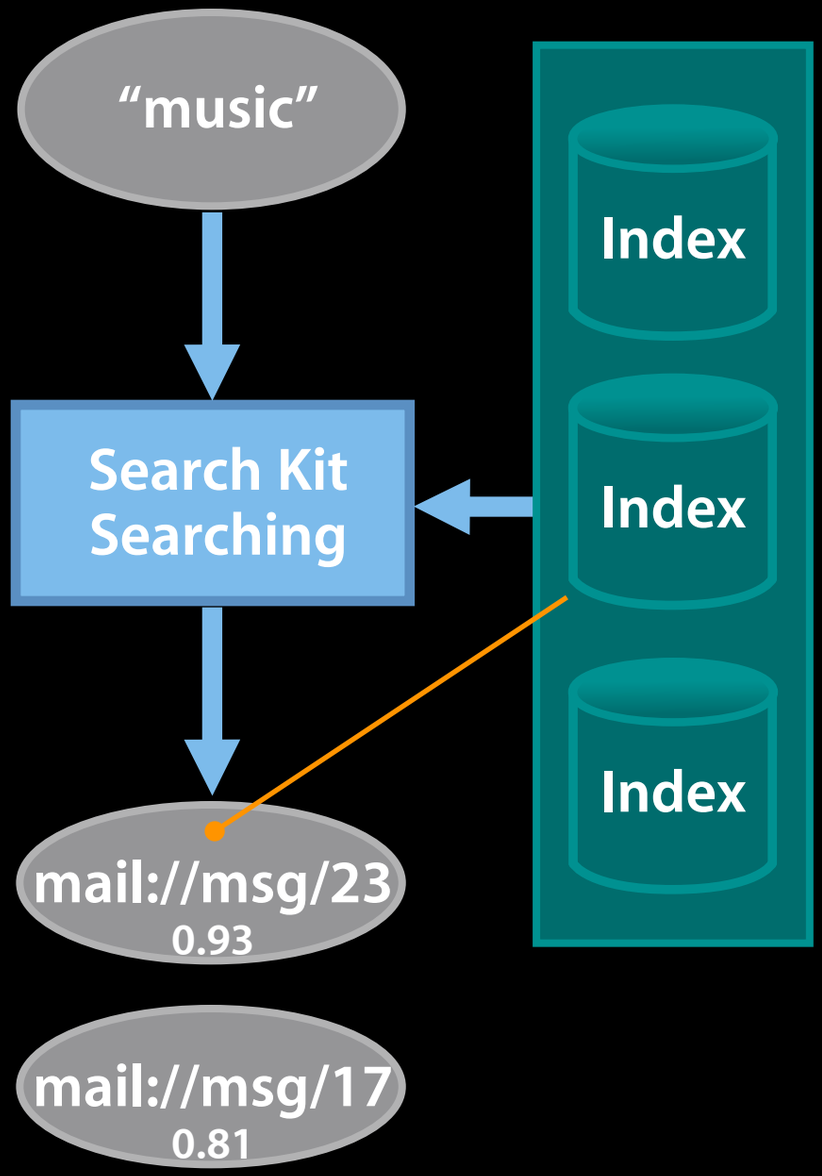
## Process results




- Search Kit returns
  - Document reference
  - Relevance ranking

# Searching

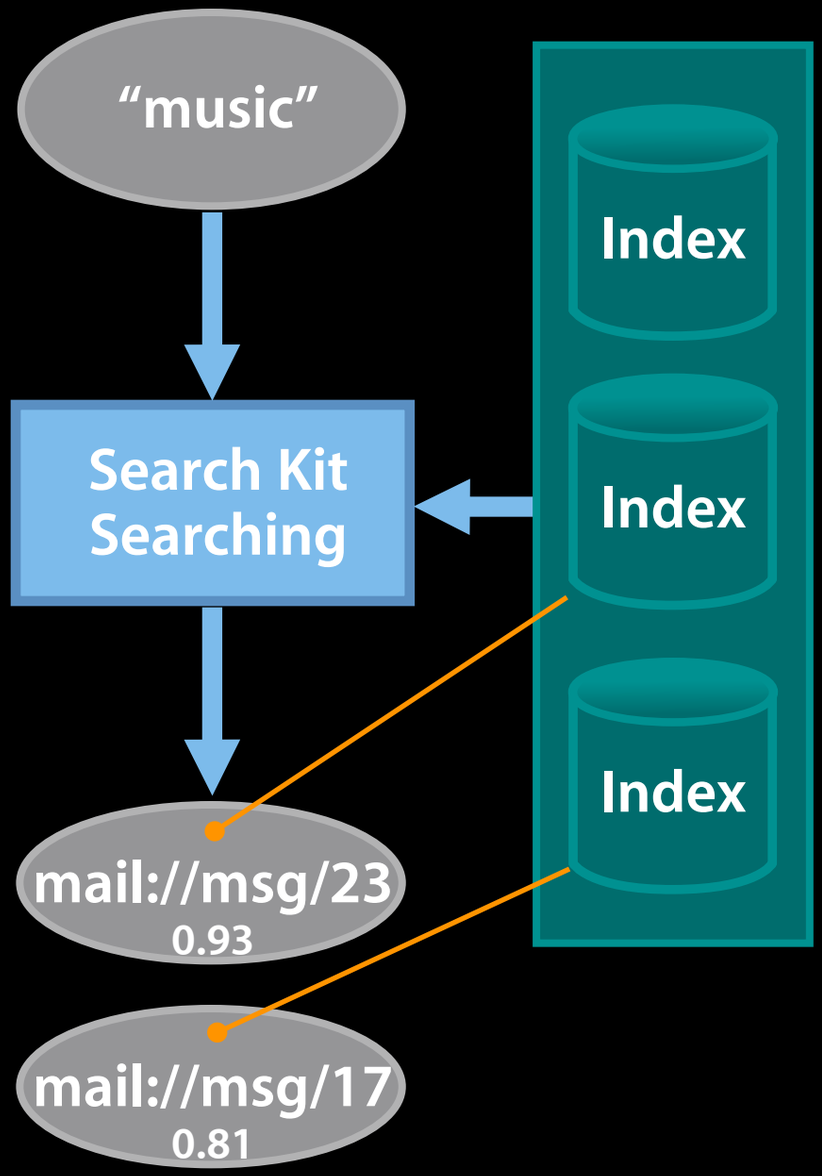
## Process results




- Search Kit returns
  - Document reference
  - Relevance ranking
  - Index reference

# Searching

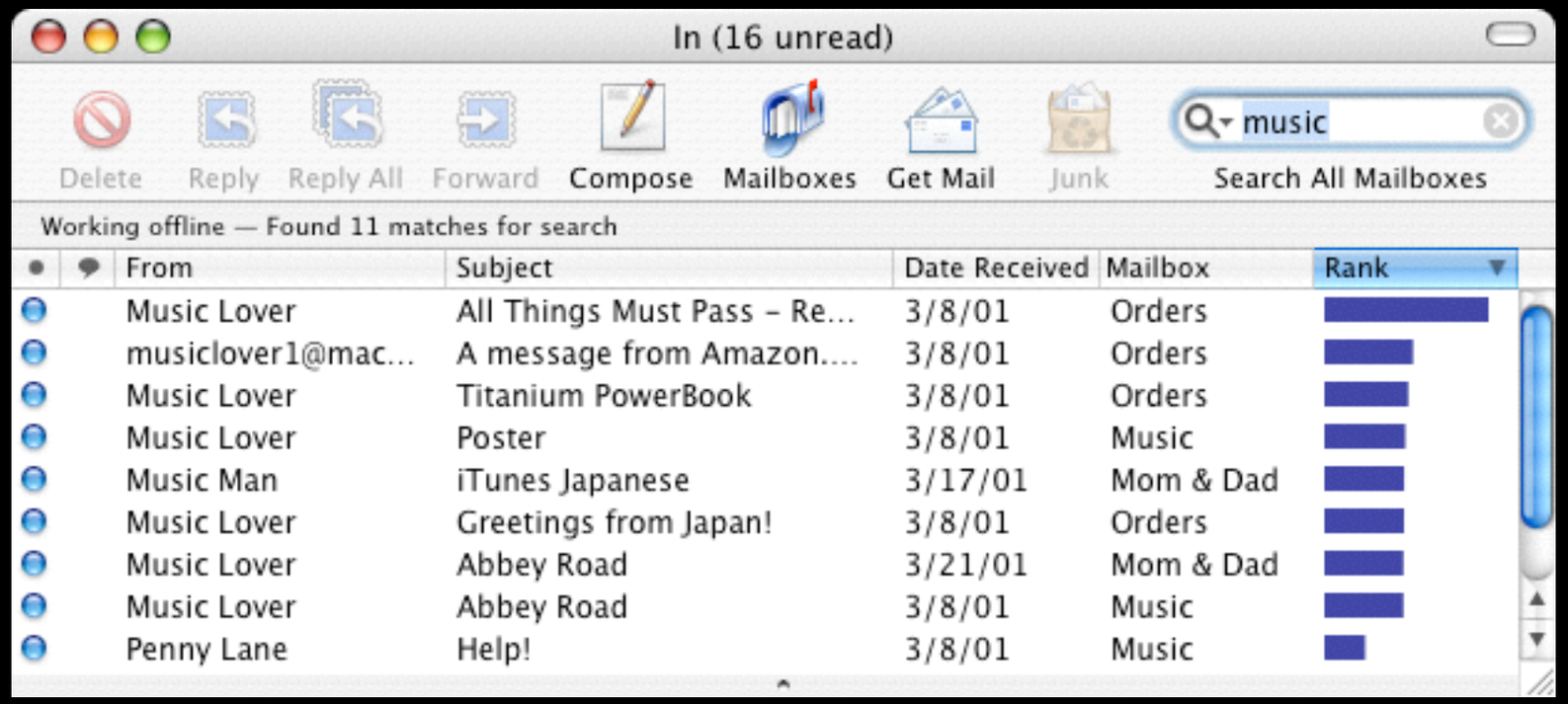
## Process results




- Search Kit returns
  - Document reference
  - Relevance ranking
  - Index reference

# Searching

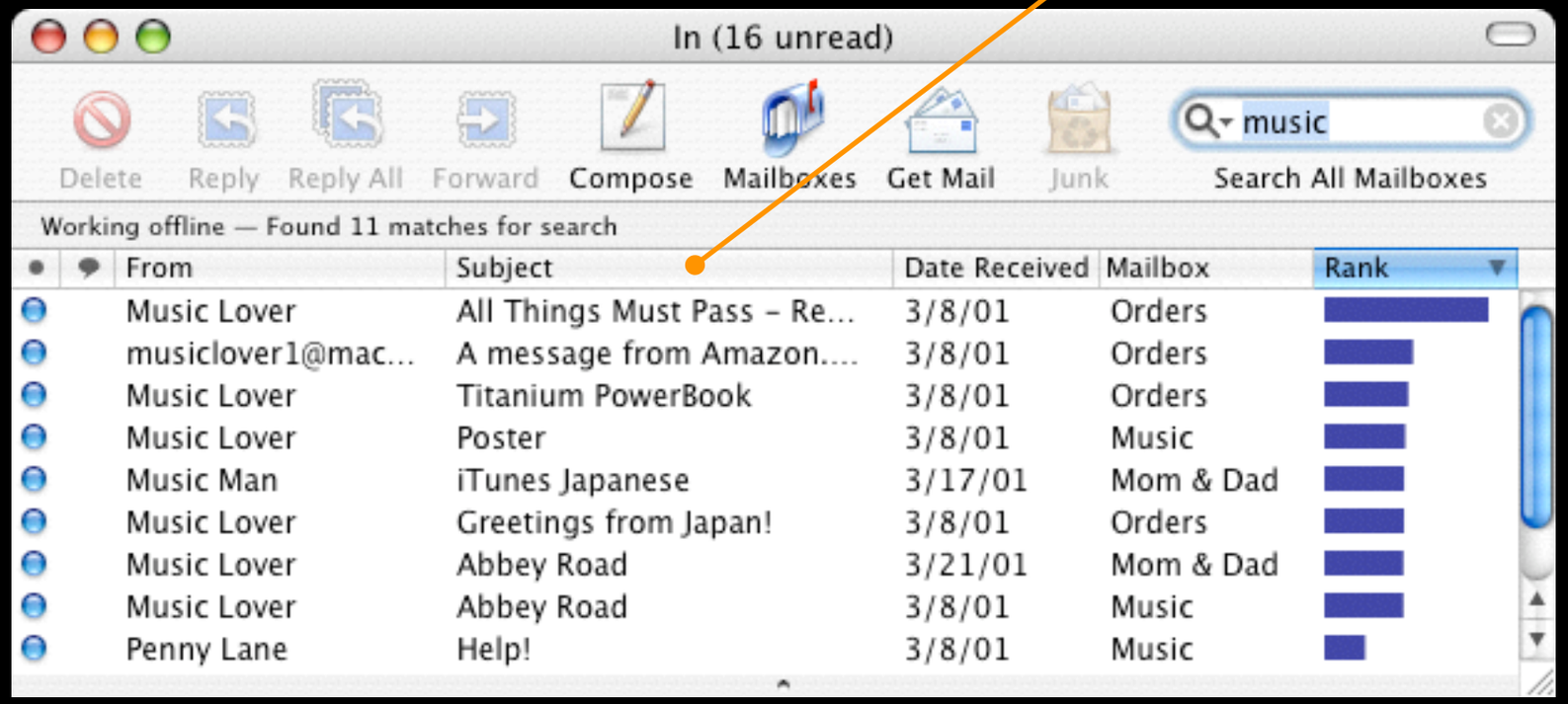
## Process results



# Searching

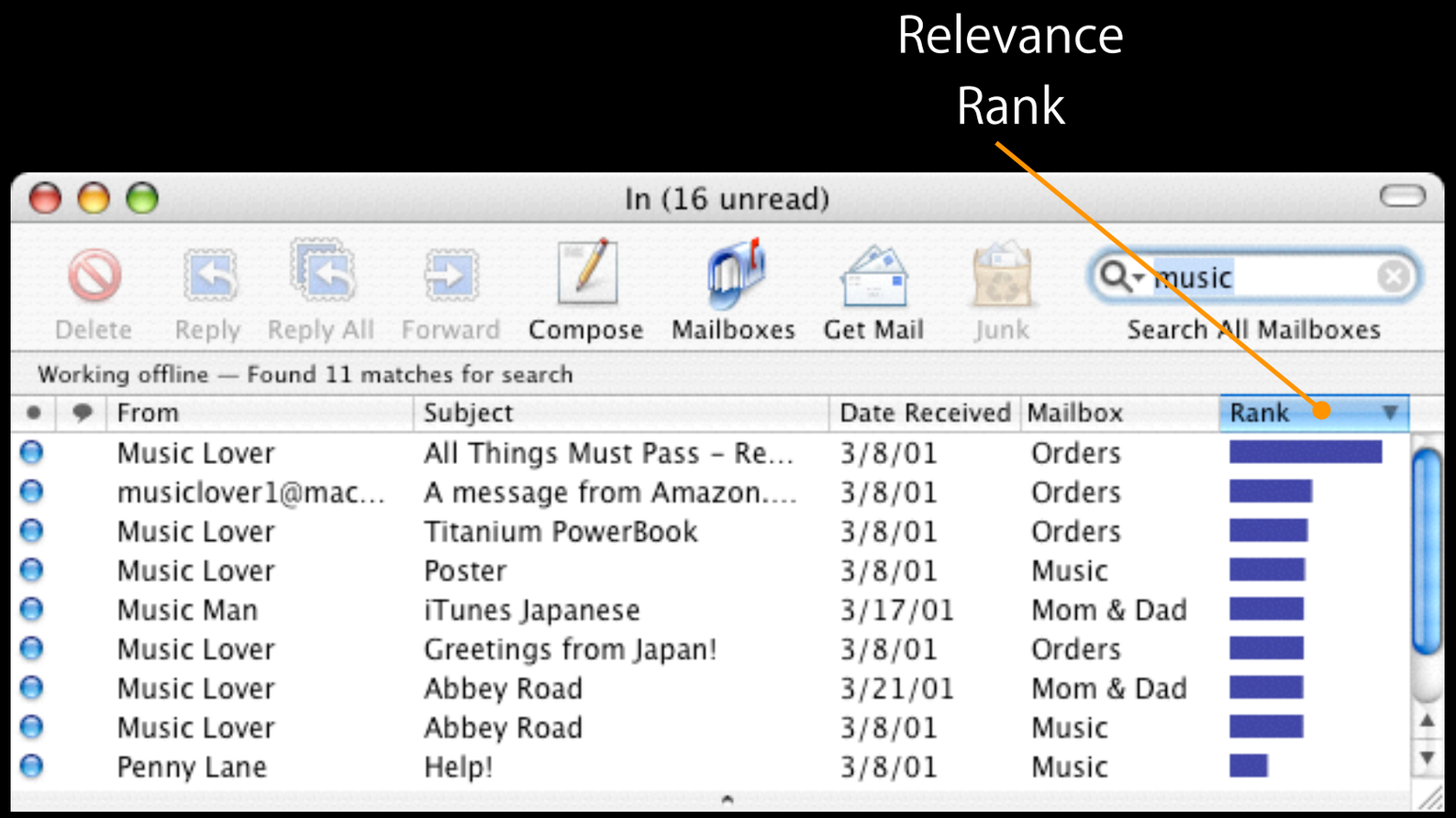
## Process results

Document
Reference



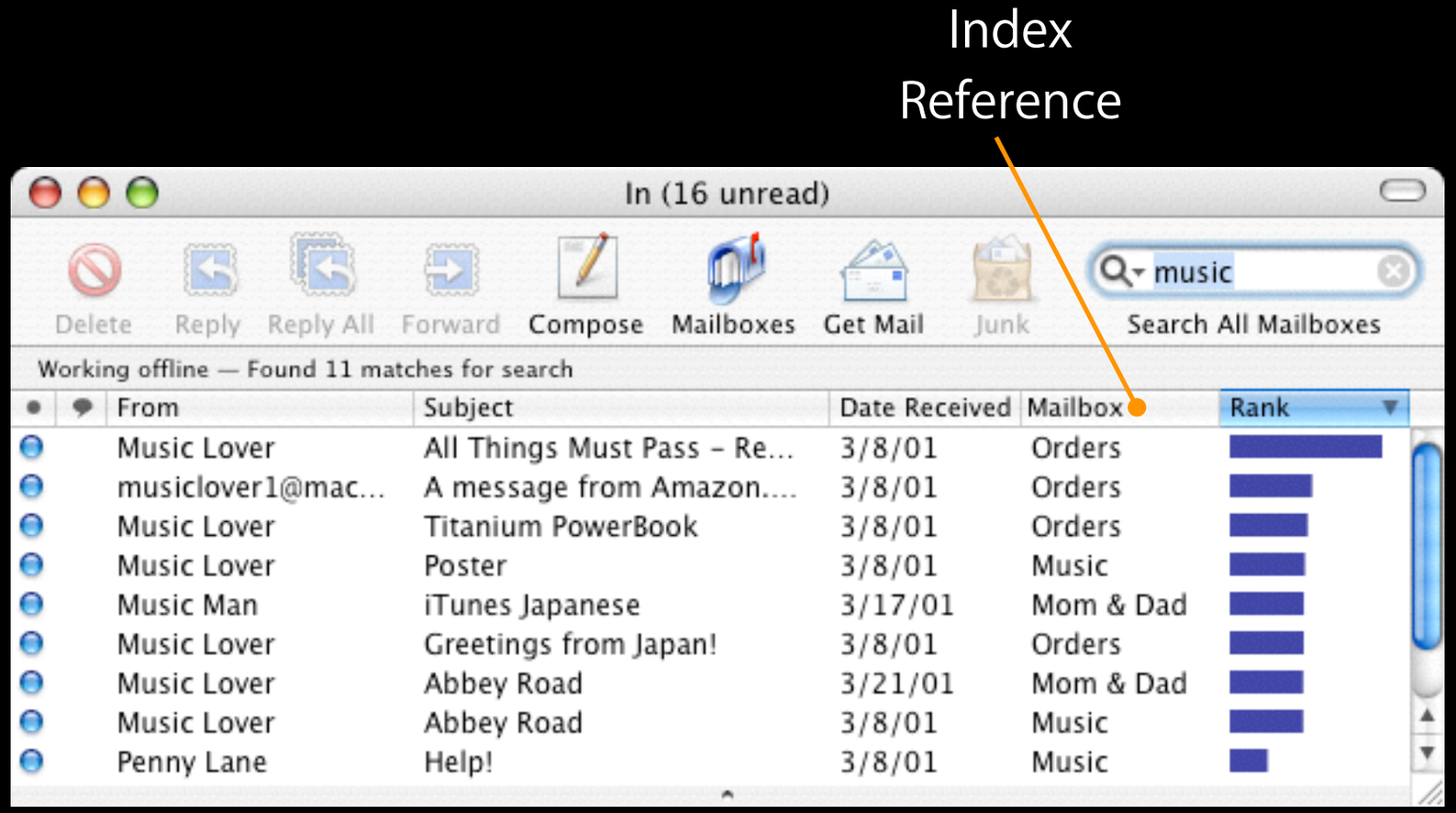
# Searching

## Process results



# Searching

## Process results



# Searching In 3 Steps

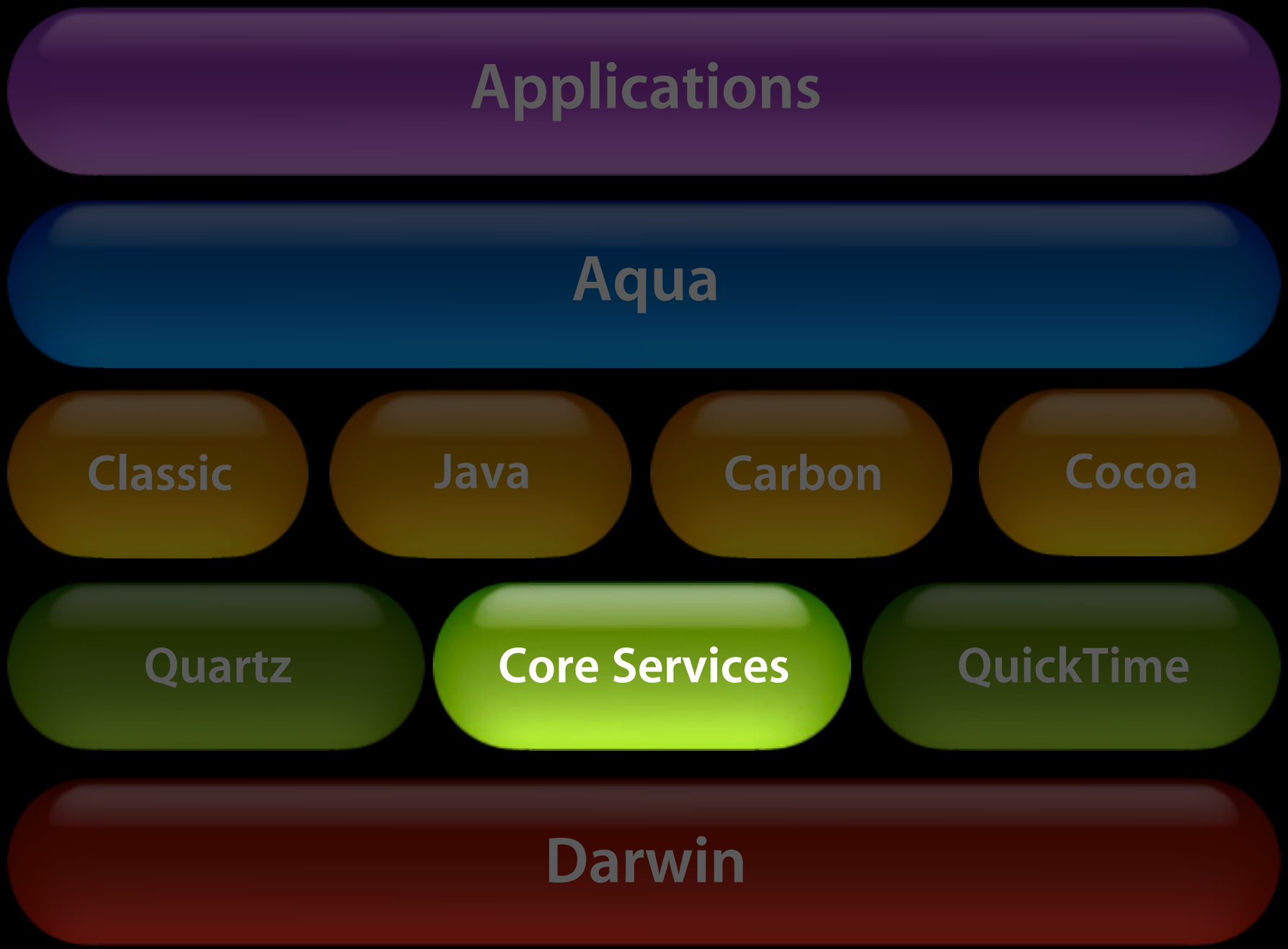
1. Create Search Group
2. Send Query
3. Process Results

## You've added searching!

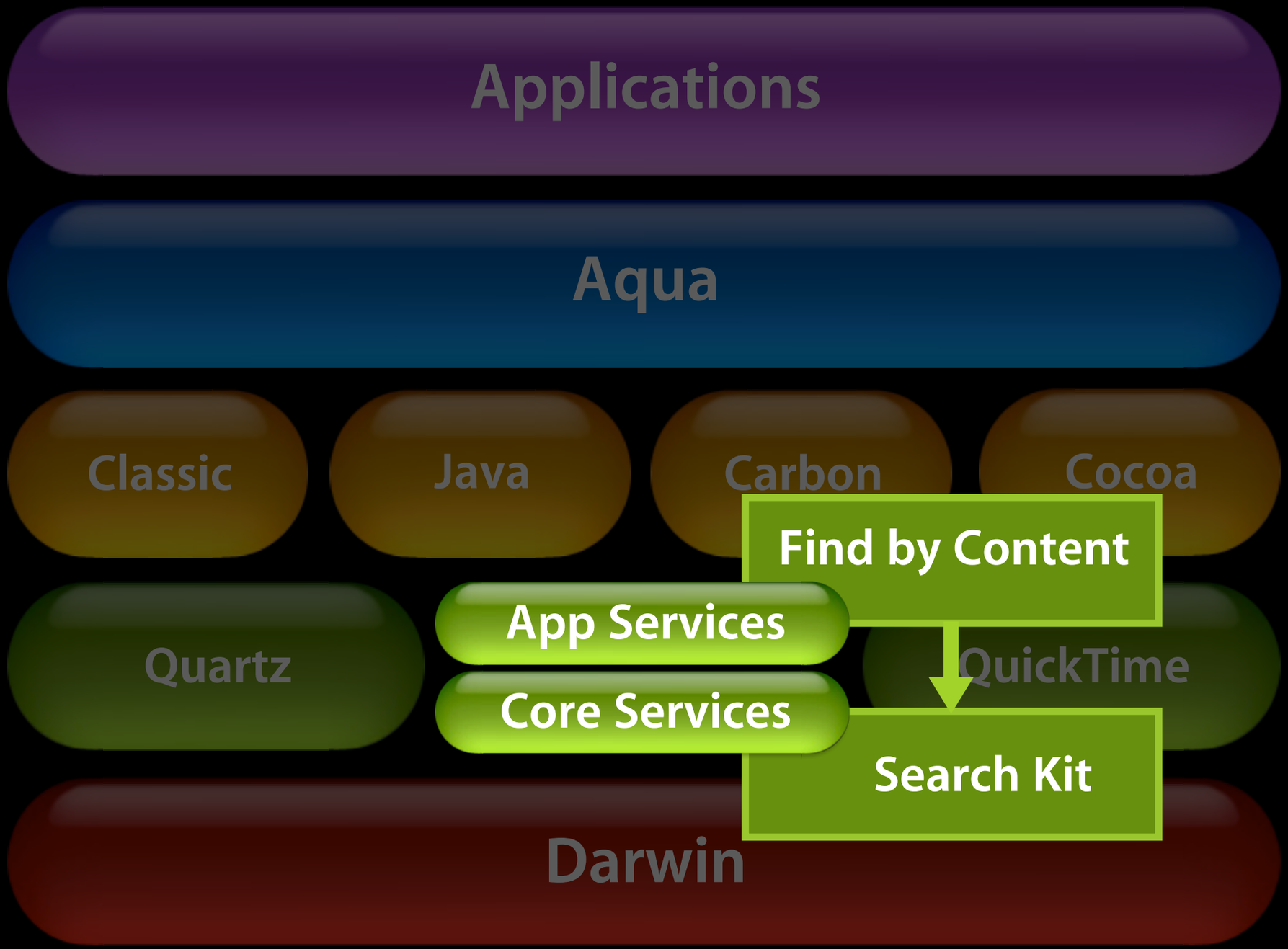
# Summarization

**David Cásseres**
Find by Content Tech Lead

# High-Level Interface to SearchKit



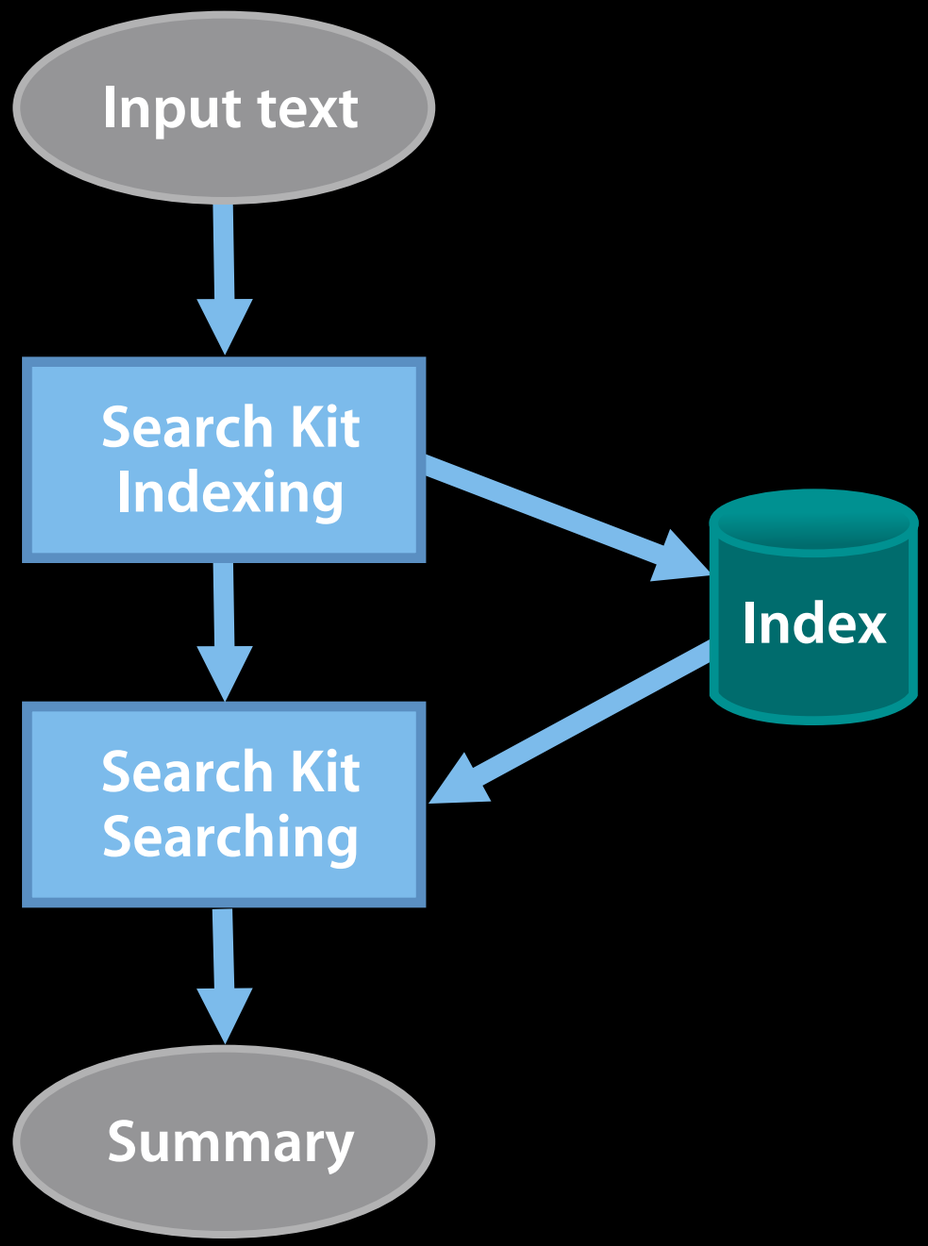
# High-Level Interface to SearchKit



# Simple Technique

## Index the sentences

- Search for the best ones



# Demo

# Why Use Summarization?

- A useful feature
- It’s fun!

# Fixed-Size Summarization

- One function does everything
  - Input CFString
  - Input desired number of sentences
  - Output CFString
- It’s that simple

# Resizable Summarization

- One function to do the hard work
- A second function to do the fast work
- Memory considerations

# How Does It Work?

- Simple statistics on words in sentences
- Finding sentences is the key
- Only complete sentences are kept
- Meaning and grammar are ignored

# Wrap Up

# Summary

- A powerful text searching framework
- Indexes anything that has text
- Provides fast searching

# Summary

- A powerful text searching framework
- Indexes anything that has text
- Provides fast searching
  - And summarization

# Summary

- A powerful text searching framework
- Indexes anything that has text
- Provides fast searching
  - And summarization

- Now the cat’s out of the bag!

# Roadmap

| | |
|---|---|
| 404 Unicode for Japanese, Chinese and Everything Else | Russian Hill<br>Tue., 2:00pm |
| 408 Apple Help Updates and How To's | Mission<br>Tue., 5:00pm |
| 417 Using the Address Book Framework | Russian Hill<br>Thur., 2:00pm |
| 418 Internet Technologies: Advanced Foundation URL APIs | Nob Hill<br>Thur., 3:30pm |

# Who to Contact

**John Geleynse**
Manager, Software Evangelism
geleynse@apple.com

http://developer.apple.com/wwdc2003/urls.html

# For More Information

# Reference Library

## Documentation

### SearchKit Reference

ADC Member Site > Download Software

### System Overview

ADC Home > Documentation > Mac OS X

# Reference Library

## Headers

### SearchKit

System > Library > Frameworks > CoreServices > Frameworks > SearchKit

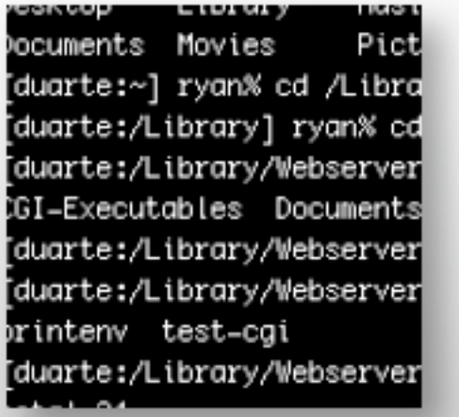
### Find by Content and Summarization

System > Library > Frameworks > ApplicationServices > Frameworks > FindByContent

### Address Book

System > Library > Frameworks > AddressBook

# Q&A




**Wayne Loofbourrow**
**David Cásseres**
**Eric Koebler, Alice Li**

http://developer.apple.com/wwdc2003/urls.html

# WWDC 2003

Worldwide Developers Conference